



Ismael Cesar da Silva Araujo

Aprendizagem de máquina quântica e comitê quântico de classificadores

Recife

2019

Ismael Cesar da Silva Araujo

Aprendizagem de máquina quântica e comitê quântico de classificadores

Monografia apresentada ao Curso de Bacharelado em Ciências da Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Ciências da Computação.

Universidade Federal Rural de Pernambuco – UFRPE
Departamento de Computação
Curso de Bacharelado em Ciências da Computação

Orientador: André Câmara Alvez do Nascimento
Coorientador: Adenilton José da Silva

Recife
2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- 183a Araujo, Ismael Cesar da Silva
Aprendizagem de máquina quântica e comitê quântico de classificadores / Ismael Cesar da Silva Araujo.
- 2019.
44 f. : il.
- Orientador: Andre Camara Alvez do Nascimento.
Coorientador: Adenilton jose da Silva.
Inclui referências.
- Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Ciência da Computação, Recife, 2019.
1. computação quântica. 2. aprendizagem de máquina quântico. 3. comitês de classificadores quânticos.
I. Nascimento, Andre Camara Alvez do, orient. II. Silva, Adenilton jose da, coorient. III. Título



MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

<http://www.bcc.ufrpe.br>

FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Trabalho defendido por ISMAEL CESAR DA SILVA ARAUJO às 14 horas do dia 02 de dezembro de 2019, no Auditório do CEAGRI-02 – Sala 07, como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado "Aprendizagem de máquina quântico e comitê quântico de classificadores", orientado por André Câmara Alves do Nascimento e aprovado pela seguinte banca examinadora:

André Câmara Alves do Nascimento
DC/UFRPE

Maigan Steffane da Silva Alcântara
DC/UFRPE

Aos meus pais, por sempre acreditarem em mim.

Agradecimentos

Agradeço a Deus por nunca ter me abandonado em toda minha trajetória. Aos meus pais e minha irmã, por terem me dado apoio. E a todos meus professores e professoras, por sua paciência, dedicação e orientação.

“O temor do Senhor é o princípio da sabedoria; bom entendimento têm todos os que cumprem os seus mandamentos; o seu louvor permanece para sempre.”
(Salmos 111:10)

Resumo

Aprendizagem de máquina quântica tratar-se de uma sub-área de computação quântica que estuda, dentre outras coisas, a criação de equivalentes quânticos de modelos clássicos de classificação. Um comitê de classificadores, trata-se de um modelo de classificação cuja saída é resultado da composição das saídas de diversos classificadores que compõe o comitê. Comitê de classificadores é um modelo de aprendizado supervisionado que pode ser utilizado como tipo de aprendizado livre de otimização de parâmetros. Ou seja, sem o uso da etapa de treinamento. Com a premissa de que ao se utilizar um conjunto suficientemente grande de classificadores medianos pode-se obter um comitê com um bom desempenho. Este trabalho investiga as diferenças de desempenho no uso equivalente quântico do comitê de classificadores, para classificadores treinados e não treinados. Onde foram feitas simulações, cujo desempenho foi mensurado a partir do cálculo de amplitudes de probabilidades do sistema. E os modelos de aprendizagem de máquina do comitê foram executados sobre bases *benchmark* disponíveis pela biblioteca *scikit-learn*.

Palavras-chave: computação quântica, aprendizagem de máquina quântico, comitês de classificadores quânticos.

Abstract

Quantum machine learning is a subarea of quantum computing that studies, among other things, the creation of equivalent classical classifiers. An ensemble of classifiers is a classification model in which the output is a combined result of the outputs of the classifiers contained in it. With the premiss that when using a sufficiently large ensemble with average classifiers, a good performance can still be obtained. This work investigates the differences in the performance of a quantum equivalent of an ensemble of classifiers, using trained and untrained classifiers. Where the simulation was made, which the performance was measured through the calculation of the amplitude probabilities of the system. And the machine learning models of the ensemble were executed over benchmark datasets made available by *scikit-learn* library.

Keywords: quantum computing, quantum machine learning, quantum ensemble of quantum classifiers.

Lista de ilustrações

Figura 1 – Esperança matemática para o caso do classificador avaliar uma saída nunca antes vista. Cujos os dados estão na distribuição normal.	22
Figura 2 – A: $\mathbb{E}[f(\tilde{x}, o, w_0)]$ com variâncias $\sigma_- = 0.4$ e $\sigma_+ = 0.9$. B: $\mathbb{E}[f(\tilde{x}, o, w_0)]$ com variâncias $\sigma_- = 0.9$ e $\sigma_+ = 0.4$	23
Figura 3 – Regiões de decisão para modelos treinados e não treinados utilizando os valores descritos de μ_{\pm} intervalos de pesos e número de modelos	24
Figura 4 – Regiões de decisão para modelos treinados não treinados utilizando os valores descritos de μ_{\pm} intervalos de pesos e número de modelos	24
Figura 5 – PPEDs para as bases de dados <i>iris</i> , <i>wine</i> e <i>breast-cancer</i> utilizando o comitê quântico com 100 classificadores não treinados	30
Figura 6 – PPEDs de <i>Hit</i> e <i>Miss</i> para as base de dados <i>iris</i> utilizando o comitê quântico com 400 classificadores treinados com 250 épocas de treinamento	33
Figura 7 – PPEDs de <i>Hit</i> e <i>Miss</i> para as base de dados <i>wine</i> utilizando o comitê quântico com 400 e 500 classificadores treinados com 500 e 250 épocas de treinamento respectivamente	35
Figura 8 – PPEDs de <i>Hit</i> e <i>Miss</i> para as base de dados <i>wine</i> utilizando o comitê quântico com 200 e 500 classificadores, para ambos os casos treinados com 500 épocas de treinamento	37

Lista de tabelas

Tabela 1 – Resultados para um comitê com 100 classificadores não treinados .	29
Tabela 2 – Resultados para um comitê com 200 classificadores não treinados .	30
Tabela 3 – Resultados para um comitê com 300 classificadores não treinados .	31
Tabela 4 – Resultados para um comitê com 400 classificadores não treinados .	31
Tabela 5 – Resultados para um comitê com 500 classificadores não treinados .	31
Tabela 6 – Resultados para um comitê de classificadores quântico de diversos tamanhos para a base de dados <i>iris</i>	32
Tabela 7 – Probabilidades por exemplo de dados média e desvio padrão para os casos de acerto (<i>Hit</i>) e de erro (<i>Miss</i>) do comitê de classificadores quântico. Para os comitês de classificadores quânticos de diversos tamanhos, com classificadores treinados com 100, 250 e 500 épocas de treinamento. Usando a base de dados <i>iris</i>	33
Tabela 8 – Resultados para um comitê de classificadores quântico de diversos tamanhos para a base de dados <i>wine</i>	34
Tabela 9 – Probabilidades por exemplo de dados média e desvio padrão para os casos de acerto (<i>Hit</i>) e de erro (<i>Miss</i>) do comitê de classificadores quântico. Para os comitês de classificadores quânticos de diversos tamanhos, com classificadores treinados com 100, 250 e 500 épocas de treinamento. Usando a base de dados <i>wine</i>	34
Tabela 10 – Resultados para um comitê de classificadores quântico de diversos tamanhos para a base de dados <i>breast-cancer</i>	36
Tabela 11 – Probabilidades por exemplo de dados média e desvio padrão para os casos de acerto (<i>Hit</i>) e de erro (<i>Miss</i>) do comitê de classificadores quântico. Para os comitês de classificadores quânticos de diversos tamanhos, com classificadores treinados com 100, 250 e 500 épocas de treinamento. Usando a base de dados <i>breast-cancer</i>	36

Lista de abreviaturas e siglas

IA	Inteligência Artificial
RNA	Redes Neurais Artificiais
AMQ	Aprendizagem de Máquina Quântico
PPED	Probabilidade Por Exemplo de Dados
MLP	<i>Multi Layer Perceptron</i>
SGD	<i>Stochastic Gradient Descent</i>

Sumário

	Lista de ilustrações	7
1	INTRODUÇÃO	11
2	COMPUTAÇÃO QUÂNTICA	16
3	COMITÊ DE CLASSIFICADORES	19
3.1	Análise de casos reduzidos de comitês de classificadores	19
3.2	Comitê quântico de classificadores	24
4	EXPERIMENTOS E RESULTADOS	27
4.1	Comitê de classificadores não treinados quântico	29
4.2	Comitê de classificadores treinados quântico	32
4.3	Discussões	37
5	CONCLUSÃO	38
	REFERÊNCIAS	40

1 Introdução

Inteligência artificial (IA) é uma sub-área da ciência da computação que estuda a criação e o uso de programas como agentes inteligentes, capazes de tomar decisões, rotular ou agrupar dados baseados na entrada recebida (Russell e Norvig 2016). Aprendizagem de máquina se trata de outra sub-área de IA que estuda principalmente as ações de rotulação e agrupamento de dados.

A tarefa de agrupar dados consiste na construção de dois ou mais agrupamentos de dados (*cluster*) onde os dados são agrupados segundo o grau de semelhança entre suas características (*features*), não se fazendo o uso de rótulos ou de algum especialista (ou professor) que possa rotular os dados previamente. Este tipo de aprendizado caracteriza o aprendizado não supervisionado. Enquanto que a tarefa de atribuir rótulos a exemplos de dados baseados em suas características, chama-se classificação.

O programa que atribui os rótulos aos exemplos de dados chama-se classificador. Geralmente um classificador mapeia vetores de alta dimensionalidade a um espaço classes, ou rótulos. Portanto, um classificador pode ser definido como uma função tal qual:

$$f : \mathcal{X} \rightarrow \mathcal{Y} \quad (1.1)$$

Onde, \mathcal{X} representa o espaço das características e \mathcal{Y} o espaço dos rótulos. No caso de classificadores parametrizados $f(x; \theta)$, os parâmetros θ são atualizados durante uma etapa afim de melhorar o desempenho do classificador na classificação de novos exemplos (Bishop 2006, Russell e Norvig 2016, Faceli et al. 2011). Ou seja, os parâmetros de um classificador são ajustados de modo que o mesmo seja capaz de generalizar um problema (Bishop 2006, Russell e Norvig 2016). Como é o caso de uma Rede Neural Artificial (RNA), a qual é um tipo de classificador cujos parâmetros são os pesos sinápticos das conexões entre neurônios de diferentes camadas (Haykin 2001). Sendo Θ o espaço dos parâmetros, na equação abaixo.

$$f : \mathcal{X}; \Theta \rightarrow \mathcal{Y} \quad (1.2)$$

Computação quântica é a área da ciência da computação a qual estuda a codificação e processamento de informação em estados quânticos (Nielsen e Chuang 2000). Para codificação e processamento de informações em um processador, computadores clássicos fazem uso de conjuntos de *bits*. Tais quais podem assumir os estados 0 ou 1 exclusivamente.

Por outro lado, computadores quânticos fazem o uso de conjuntos de bits quânticos (*qubits*), onde cada *qubit* tem a propriedade de assumir o estado de 0, 1, ou 0 e 1

ao mesmo tempo. Esta propriedade ímpar de um *qubit* permite a um computador quântico representar um número exponencialmente grande de estados ao mesmo tempo, de acordo com o número de *qubits* no sistema. A propriedade de um *qubits* de assumir mais de um estado ao mesmo tempo é chamada de superposição.

E assim como em computadores clássicos, computadores quânticos podem aplicar operações nos estados dos *qubits*. Essas operações agem de forma linear sobre os estados, ou seja, em computadores quânticos as operações podem ser aplicadas sobre diversos estados ao mesmo tempo, ação conhecida como paralelismo quântico. Essas propriedades representam uma vantagem não só em relação a redução de custo computacional, como também em relação a representação de sistemas e fenômenos que até então são inviáveis de simular em um computador clássico. (Feynman 1982).

Como dito anteriormente, o processamento de dados em aprendizagem de máquina geralmente envolve efetuar operações sobre vetores de alta dimensionalidade. Isso é verdade principalmente para as RNAs (Bishop 2006, Haykin 2001, Faceli et al. 2011). As propriedades ímpares que um sistema quântico com múltiplos *qubits* possui podem ser aproveitadas não só para acelerar como para aprimorar algoritmos de aprendizado.

O ramo de pesquisa que explora esses potenciais chama-se Aprendizagem de Máquina Quântico (AMQ) (Biamonte et al. 2017). Diferentes abordagens de aprendizagem de máquina estudam como transformar um algoritmo de aprendizado clássico em uma versão quântica análoga, um processo conhecido como quantização.

Outras abordagens estudam a criação de classificadores especificamente voltados para sistemas quânticos. Mas seja por quantização ou através de um classificador específico, acredita-se que um algoritmo de aprendizado quântico seja capaz de reconhecer padrões que algoritmos clássicos não conseguiriam. Isso devido ao poder exponencialmente grande de representação de estados de um sistema quântico (Biamonte et al. 2017).

Outros trabalhos focam em como os dados podem ser codificados em um sistema quântico para que os mesmos sejam processados. Existindo a possibilidade das características serem codificadas nas amplitudes de probabilidade do sistema (*phase encoding*), ou os dados serem integralmente inscritos num estado em superposição em representação binária (*qubit encoding*) (Grant et al. 2018, Ventura 1998, Trugenberger 2002, Schuld, Fingerhuth e Petruccione 2017). Possibilitando a utilização dos mesmos tipos de abstrações relacionadas a ponto flutuante ou complemento a dois.

Contudo a principal limitação nesta área continua sendo o número de *qubits*. Pois quanto maior o número de *qubits* utilizados no sistema, mais o mesmo se torna estável dificultando a tarefa de manter um estado em superposição por tempo suficiente

para que os cálculos sejam terminados.

Quando um problema de classificação é muito complexo, pode-se fazer o uso de um conjunto de classificadores. Utilizando um conjunto suficientemente grande de classificadores medianos pode efetuar boas classificações. Neste trabalho, classificadores medianos são classificadores cujo desempenho é pouco melhor que rotulação aleatória. Mesmo que esse conjunto de classificadores tenham um desempenho mediano, considerando a classificação de todo o conjunto é possível efetuar boas classificações (Schapire 1990, Kearns e Valiant 1988).

Seja f um classificador fraco qualquer, se f tiver um desempenho um pouco melhor que atribuição de classe aleatória o mesmo tem uma boa probabilidade de classificar uma fração do conjunto de dados corretamente (Schapire 1990, Kearns e Valiant 1988). Quando a classificação é efetuada f é utilizado para avaliar uma entrada \tilde{x} sob o parâmetro θ_x . Cujas classe \tilde{y} é atribuída a entrada por $\tilde{y} = f(\tilde{x}, \theta)$.

Para melhor compreensão pode-se adotar como analogia o fato de que quando alguém tem uma decisão importante a fazer. Uma decisão sobre a qual não se tem nenhum conhecimento, e que é necessário consultar um grupo de especialistas. Avalia-se os conselhos de cada especialista de acordo com algum critério, e então toma-se a decisão. O conjunto de especialistas aqui representam os classificadores. E o conjunto de classificadores corresponde a um comitê.

Dado o conjunto de classes $y = \{-1, 1\}$, um comitê de classificadores avalia uma nova entrada segundo o conjunto de parâmetros, e para evitar que classificações erradas influenciem no processo de classificação, pesos são atribuídos cada um dos parâmetros. Tais pesos determinam o grau de influência que cada classificador terá na decisão.

$$\tilde{y} = \text{sgn} \left(\sum_{\theta} w_{\theta} f(\tilde{x}; \theta) \right) \quad (1.3)$$

Onde *sign* trata-se da função sinal.

$$f(x) = \begin{cases} -1, & \text{se } x < 0 \\ 1, & \text{se } x \geq 0 \end{cases} \quad (1.4)$$

Retomando a analogia feita anteriormente, e se fosse possível consultar todos os especialistas ao mesmo tempo? A avaliação dos classificadores pode ser feita segundo um número exponencialmente grande de parâmetros tirando vantagem da propriedade de superposição. Porém o resultado seria obtido com um certo grau de probabilidade.

Em (Schuld e Petruccione 2018) foi investigado o uso de comitês de classificadores quânticos não treinados. A decisão (ou rotulação) de cada classificador era ponderada pela acurácia do mesmo. Contudo o uso de comitês de classificadores tais quais os descritos em (Schuld e Petruccione 2018) dependiam de uma boa distribuição dos dados.

Para a solução do problema da dependência do desempenho dos classificadores do comitê quântico em relação distribuição dos dados, neste trabalho é estudado a inclusão de uma etapa de treinamento para os classificadores do comitê.

Treinamento de classificadores quânticos é um tema que está sendo estudado na área de computação quântica, com métodos como amplificação de amplitude para efetuar treinamento e meta-treinamento de RNAs quânticas (Fawaz et al. 2019). Neste trabalho, entretanto, será levada em consideração uma abordagem treinamento dos classificadores (Santos et al. 2018, Santos et al. 2018) onde memórias associativas quânticas são utilizadas para codificação dos dados em estados quânticos, porém utilizando uma abordagem mais simples de validação cruzada, *holdout*.

Os principais objetivos deste trabalho são:

- **Principal:** Executar uma simulação de comitês quânticos de classificadores quânticos treinados e não treinados sobre bases de dados *benchmark*.
- **Secundário:** A partir dos resultados das simulações destacar a diferença de desempenho de comitês quânticos de classificadores.

Onde a contribuição deste trabalho em relação ao uso de comitês quânticos de classificadores quânticos é a adição da etapa de treinamento na construção do comitê.

Os resultados foram obtidos a partir de simulações de casos reduzidos utilizando bases de dados *benchmark* disponíveis pela biblioteca *scikit-learn* (Pedregosa et al. 2011) usando a linguagem de programação *python*. Onde as bases de dados utilizadas foram *iris*, *breast cancer* e *wine*. E o desempenho do comitê foi mensurado de acordo com a probabilidade do mesmo de retornar a classe correta para cada exemplo do conjunto de teste. Onde a probabilidade do sistema foi utilizada como parâmetro de medição pois a saída de um sistema quântico é probabilística, de modo que sempre que um sistema está em superposição e uma operação de medição é aplicada o estado entra em colapso e um valor bem definido com um certo grau de probabilidade.

Nesta seção foram apresentados alguns conceitos introdutórios e problemáticas em relação não só ao uso de computação quântica, como também ao uso de computação quântica a aprendizagem de máquina e comitês quânticos de classificadores. O resto deste trabalho está estruturado da seguinte forma: No capítulo 2 são apresentados os principais conceitos relacionados a computação quântica. No capítulo 3 é

apresentado o conceito de comitês de classificadores quânticos, bem como são apresentados algumas réplicas dos resultados apresentados em (Schuld e Petruccione 2018). No capítulo 4 é apresentada a descrição dos experimentos feitos, também quais foram os resultados obtidos e discussões sobre os resultados. No capítulo 5 são apresentadas as considerações finais, bem como são apresentados possíveis trabalhos futuros.

2 Computação quântica

Um *qubit* é a menor unidade de base computacional utilizada por computadores quânticos para processamento de informação (Nielsen e Chuang 2000). O estado de um *qubit* pode ser representado matematicamente com o segundo a notação vetor coluna. Porém por simplicidade adota-se a notação de Dirac (McMahon 2007, Yanofsky, Mannucci e Mannucci 2008) (2.1).

$$|0\rangle = \begin{bmatrix} 1 & 0 \end{bmatrix}^T; |1\rangle = \begin{bmatrix} 0 & 1 \end{bmatrix}^T \quad (2.1)$$

Seja $|\psi\rangle$ o estado arbitrário de um *qubit* na equação (2.2), o estado do *qubit* apresentado encontra-se em superposição. Superposição é a propriedade de um *qubit* de assumir dois estados ao mesmo tempo, no caso da equação (2.2) o *qubit* encontra-se nos estados $|0\rangle$ e $|1\rangle$ ao mesmo tempo. Para se obter alguma informação útil desse estado é necessário que seja efetuada uma operação de medição. No entanto, não é possível obter todos os estados de um *qubit* em superposição sem que o resto da informação contida no mesmo seja afetada, isto também vale para estados quânticos com múltiplos *qubits* em superposição. Pois, ao se efetuar uma medição a superposição entra em colapso e o *qubit* assume um estado específico com um grau de probabilidade. Em (2.2) $|\psi\rangle$ assumirá o estado $|0\rangle$ com probabilidade $|\alpha|^2$ e $|1\rangle$ com probabilidade $|\beta|^2$ após a medição.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (2.2)$$

Onde os escalares α e β são números complexos que determinam a amplitude de probabilidade de cada estado. Um estado também pode conter múltiplos *qubits*. Seja em notação de Dirac ou em notação de vetores coluna, um estado com múltiplos *qubits* é representado efetuando produtos tensoriais entre dois estados desejados da base. Como por exemplo em notação de Dirac: $|0\rangle \otimes |1\rangle = |01\rangle$. Arbitrariamente, um estado quântico com $|x\rangle \in \mathbb{C}^N$ e com todos os estados em superposição pode ser representado como na equação (2.3) Onde N representa o número de *qubits* no sistema.

$$|\psi\rangle = \sum_{x \in \{0,1\}^N} \alpha_x |x\rangle \quad (2.3)$$

Foi mencionado anteriormente que um *qubit* pode ser representado utilizando a notação vetores coluna. O fato é que esses vetores formam uma base ortonormal no

espaço vetorial \mathbb{C}^N , com N variando no intervalo $[2, 2^n]$. Uma base ortonormal trata-se de um conjunto de vetores cujo produto interno entre eles é 0, ou seja, os vetores do conjunto são ortogonais entre si e a norma (ou a magnitude) de todos os vetores do conjunto é 1.

Sendo assim, seja $|x\rangle \in \mathbb{C}^2$ e $x = \{0, 1\}^n$, com $|x\rangle$ pertencente à base ortonormal do espaço vetorial \mathbb{C}^2 . Digamos que o estado na equação (2.3) seja um sistema quântico de oito *qubits*. E considere o fato de que deseja-se medir os dois últimos *qubits* do sistema. Seja a expressão regular $\bar{x} \in \{\{0, 1\}^{N-2} + 01\}$, uma expressão que indica todas as cadeias binárias de tamanho N que terminam em 01, na equação (2.4). O sistema retornará os dois últimos *qubits* no estado 01 com probabilidade $\sum_{\bar{x}} |\alpha_{\bar{x}}|^2$. Então o estado $|\psi\rangle$ entraria em colapso após a medição, tornando-se $|\psi'\rangle$. Onde o denominador $\sqrt{\sum_{\bar{x}} |\alpha_{\bar{x}}|^2}$ serve como fator de normalização do estado resultante.

$$|\psi'\rangle = \sum_{\bar{x} \in \{0,1\}^{N-2} + 01} \frac{\alpha_{\bar{x}} |\bar{x}\rangle}{\sqrt{\sum_{\bar{x}} |\alpha_{\bar{x}}|^2}} \quad (2.4)$$

Aplicar computações em sistemas quânticos consiste em manipular não só os estados dos *qubits* como também a amplitude de probabilidade dos estados. Para efetuar essas manipulações faz-se o uso de operadores quânticos ou portas quânticas. Matematicamente, os operadores quânticos são representados como matrizes unitárias $M \in \mathbb{C}^{N \times N}$. Sendo a unitariedade dos operadores necessária devido a propriedades intrínsecas da física quântica (Yanofsky, Mannucci e Mannucci 2008). Analogamente, portas quânticas de múltiplos *qubits* podem ser combinadas através do produto tensorial para que operem sobre múltiplos *qubits*. Em Eq.(2.5) é aplicada o equivalente quântico da porta **not** no segundo *qubit*. Onde o operador identidade I significa que nenhuma operação é efetuada nos outros *qubits*.

$$I \otimes X \otimes I |0\rangle|0\rangle|0\rangle = (IXI)|0\rangle|0\rangle|0\rangle = |0\rangle|1\rangle|0\rangle \quad (2.5)$$

Nem todos os operadores quânticos podem ser decompostos em um produto tensorial de um conjunto de operadores simples (McMahon 2007). Um caso especial de operadores são aqueles que condicionalmente efetuam uma mudança do estado do sistema. Ou seja de acordo com um conjunto de um ou mais *qubits* de controle que assumirem um estado específico os mesmos operadores aplicarão a mudança desejada no estado. Esse fato é especialmente importante quando se está lidando com um estado com múltiplos *qubits* em superposição. Permitindo condicionalmente mudar os estados do sistema sem que haja necessidade de medição. Na equação (2.6) há um exemplo de uma aplicação do operador condicional *Toffoli*, ou *Tf*. Onde

\oplus e \wedge representam as operações lógicas *xor* (ou-exclusivo) e *and*, respectivamente.

$$Tf|x\rangle|y\rangle|1\rangle \rightarrow |x\rangle|y\rangle|1 \oplus (x \wedge y)\rangle \quad (2.6)$$

Uma consequência vinda do uso do operador *Toffoli* em computadores quânticos como na equação (2.6) ocorre devido ao fato de que este tipo de aplicação é equivalente a um *NAND* lógico em um computador clássico (2.7). Onde a porta lógica *NAND* é considerada uma porta universal em circuitos digitais, devido ao fato que combinando um conjunto de portas *NAND* pode se implementar qualquer outra porta lógica (Tocci e Widmer 2002). Consequentemente é possível simular qualquer função binária em um computador quântico a partir de combinações de portas *Toffoli*.

$$Tf|x\rangle|y\rangle|1\rangle \rightarrow |x\rangle|y\rangle|\neg(x \wedge y)\rangle \quad (2.7)$$

Operadores quânticos podem ser aplicados de forma linear no sistema. Isso significa que utilizando um computador quântico, é possível aplicar o operador quântico sobre um número exponencialmente grande de estados ao mesmo tempo com apenas uma aplicação. Este tipo de aplicação de operadores linearmente nos estados trata-se de paralelismos quântico. Sendo uma das propriedades que incentivou estudos relacionados a o uso de computação quântica em aprendizagem de máquina.

3 Comitê de classificadores

Como dito anteriormente, um classificador parametrizado pode ser descrito como um mapeamento $f : X, \Theta \rightarrow Y$. Onde Θ trata-se do espaço dos parâmetros que descrevem o classificador. Porém, um valor $\theta \in \Theta$ será referido como classificador em si, já que é a partir da otimização dos parâmetros que o classificador pode melhorar a classificação. Sendo assim um comitê de classificadores trata-se de um conjunto de classificadores $\{\theta_0, \dots, \theta_{n-1}\}$ cuja decisão é combinada para rotular um exemplo de dados \tilde{x} . Desse modo a saída de classificadores que reconhecem padrões diferentes podem ser utilizada para se obter a rotulação. Considerando um caso de classificação binária em que $Y = \{-1, 1\}$.

A saída dos classificadores pode ser combinada como uma soma ponderada, onde os pesos indicam o grau de influência desejado para cada classificador(3.1). Um conjunto de classificadores suficientemente grande pode ter um bom desempenho na rotulação de exemplos de dados mesmo se forem utilizados classificadores medianos. Uma opção de fator de ponderamento na utilização de um comitê trata-se da acurácia. Desse modo classificadores com baixa acurácia teriam pouca influência no momento da rotulação (Schuld e Petruccione 2018).

$$\tilde{y} = \text{sign} \left(\sum_{\theta} a_{\theta} f(\tilde{x}, \theta) \right) \quad (3.1)$$

Onde sign trata-se da função sinal e a_{θ} a acurácia correspondente ao modelo θ . Levando em conta métodos como validação cruzada, os valores de acurácia a_{θ} podem ser obtidos a partir de uma etapa adicional. Onde para um conjunto de dados D os valores de acurácia podem ser pré-computados a partir de exemplos de dados no conjunto de treinamento. Sendo esta pré-computação uma etapa necessária caso deseja-se utilizar um comitê de classificadores cujas saídas são ponderadas pelas acurácias dos mesmos.

Baseado na noção de um comitê de classificadores clássico pode-se criar um equivalente quântico. Porém antes de ponderar sobre o passo a passo da construção de um comitê de classificadores quânticos é importante entender o quão bem um comitê de classificadores medianos funciona.

3.1 Análise de casos reduzidos de comitês de classificadores

Em (Schuld e Petruccione 2018) é sugerido o uso de comitês de classificadores como um derivado de aprendizado Bayesiano, onde a classificação é independente de

otimização de parâmetros (Mitchell 1997). O problema ao se utilizar comitês de classificadores desta forma é a tarefa de classificação torna-se dependente da distribuição dos dados (Schuld e Petruccione 2018) e pelo fato de ser livre de otimização os valores aos quais os parâmetros dos classificadores são inicializados também influenciam na decisão. Os gráficos a seguir foram resultados de réplicas de análises numéricas de exemplos com casos reduzidos (ou exemplos *toy*) para comitês de classificadores clássicos utilizando casos unidimensionais e bidimensionais.

Para o caso unidimensional, com um classificador binário $f(x, o, w_0) = \text{sign}(o(x - w_0))x$ onde w_0 define a fronteira de decisão e o a orientação. Por conveniência será adotado orientação com valor $o = 1$. O classificador f define duas regiões de decisão na reta dos valores reais pertencentes a uma das classes $Y = \{-1, 1\}$. Regiões as quais são definidas como $\mathcal{R}_{-1} = [-\infty, w_0]$ e $\mathcal{R}_1 [w_0, \infty]$.

Seja o par (x, y) variáveis aleatórias e considerando o fato que a densidade de probabilidade do conjunto $p(x, y)$ de dados já é conhecida. Computar acurácia (3.2) de um ponto de vista contínuo consistiria em computar a integral (Schuld e Petruccione 2018):

$$a(\theta) = \frac{1}{2} \int_{\mathcal{R}_{-1}} p(x, y = -1) dx + \frac{1}{2} \int_{\mathcal{R}_1} p(x, y = 1) dx \quad (3.2)$$

Agora seja a função densidade de probabilidade definida como $p(x, y = \pm 1) = \rho(x; \mu_{\pm}; \sigma_{\pm}) = p_{\pm}(x)$ tal que $\int_{-\infty}^{+\infty} p_{\pm}(x) = 1$. Para uma função densidade ρ tanto para as classes $Y = \{-1, 1\}$. Cujos parâmetros são o exemplo de entrada x , média μ_{\pm} e variância σ_{\pm} da distribuição do conjunto de dados D . Sendo o fator $\frac{1}{2}$ necessário para que o valor da acurácia na equação (3.2) necessário para que o mesmo permaneça no intervalo $[0, 1]$

Desse modo ao se fazer análise numérica para um classificador unidimensional, tem-se por objetivo calcular a esperança matemática da classificação como proporcional à integral do classificador da acurácia multiplicada pela saída do classificador (Schuld e Petruccione 2018).

$$\mathbb{E}[f(\tilde{x}, o, w_0)] \propto \int a(\theta) d\theta f(\tilde{x}, o, w_0) \quad (3.3)$$

Para uma entrada nunca processada \tilde{x} . Onde para se computar os valores resultantes à integral em relação a acurácia certos fatos foram assumidos como verdade. Por exemplo, sendo $P_{\pm}(x)$ a integral da função densidade $p_{\pm}(x)$, tal que se $x \mapsto -\infty$ então $P_{\pm}(x) \mapsto 0$ e se $x \mapsto +\infty$ então $P_{\pm}(x) \mapsto 1$. Assim como havia sido descrito de forma semelhante em (Schuld e Petruccione 2018). De modo que as seguintes

propriedades podem ser obtidas a partir destas definições:

$$\int_{-\infty}^a p_{\pm}(x)dx = P_{\pm}(a) - P_{\pm}(-\infty) = P_{\pm}(a) \quad (3.4)$$

$$\int_a^{+\infty} p_{\pm}(x)dx = P_{\pm}(+\infty) - P_{\pm}(a) = 1 - P_{\pm}(a)$$

Sabendo que w_0 trata-se do parâmetro que define a fronteira de decisão neste exemplo unidimensional, o mesmo pode ser utilizado para efetuar a integração na equação (3.2). Contudo, o cálculo do valor de $a(\theta)$ deve ser feito também se levando em conta a orientação $o = \{-1, 1\}$ das regiões de decisão. Sendo assim, para o exemplo unidimensional, $a(\theta)$ deve ser calculada para caso se a orientação das regiões for $o = 1$ ou se $o = -1$. Onda a função $\text{sign}(o(\tilde{x} - w_0))$ é omitida por simplicidade de visualização. Aplicando-se as definições apresentadas, w_0 integrações e avaliando o termo $a(\theta)$ da equação (3.4) obtém-se a expressão a seguir:

$$\left(\frac{1}{2} \int_{-\infty}^{w_0} p_{-}(x)dx + \frac{1}{2} \int_{w_0}^{+\infty} p_{+}(x)dx \right) + \left(\frac{1}{2} \int_{w_0}^{+\infty} p_{+}(x)dx + \frac{1}{2} \int_{-\infty}^{w_0} p_{-}(x)dx \right) \quad (3.5)$$

Onde o primeiro e o segundo termo da soma dos elementos entre parênteses trata-se da aplicação de $a(\theta)$ segundo as orientações das regiões de decisão $o = 1$ e $o = -1$. Porém a mesma constaria multiplicando o primeiro termo entre parênteses com orientação $o = 1$ e o segundo termo com orientação $o = -1$. Neste caso unidimensional faz-se o uso da orientação da reta para definir de qual lado da reta real cada região de decisão pertence caso $o = 1$ ou $o = -1$. Desenvolvendo os termos na equação (3.5) obtêm-se:

$$\left(\frac{1}{2} [P_{-}(w_0) - P_{-}(-\infty)] + \frac{1}{2} [P_{+}(+\infty) - P_{+}(w_0)] \right) + \quad (3.6)$$

$$\left(\frac{1}{2} [P_{+}(+\infty) - P_{+}(w_0)] + \frac{1}{2} [P_{-}(w_0) - P_{-}(-\infty)] \right)$$

Aplicando as definições na equação (3.6) obtém-se a seguinte expressão:

$$\left(\frac{1}{2} P_{-}(w_0) - \frac{1}{2} P_{+}(w_0) + \frac{1}{2} \right) + \left(\frac{1}{2} P_{-}(w_0) - \frac{1}{2} P_{+}(w_0) + \frac{1}{2} \right) \quad (3.7)$$

Avaliando a expressão (3.7) e aplicando-a no segundo termo da expressão (3.3) obtém-se a equação (3.8).

$$\int_{-\infty}^{+\infty} (P_{-}(w_0) - P_{+}(w_0)) \text{sign}(\tilde{x} - w_0) dw_0 \quad (3.8)$$

Em (Schuld e Petruccione 2018) a função distribuição de probabilidade utilizada foi a distribuição gaussiana cujos hiper parâmetros para as classes $\{-1, 1\}$ utilizados

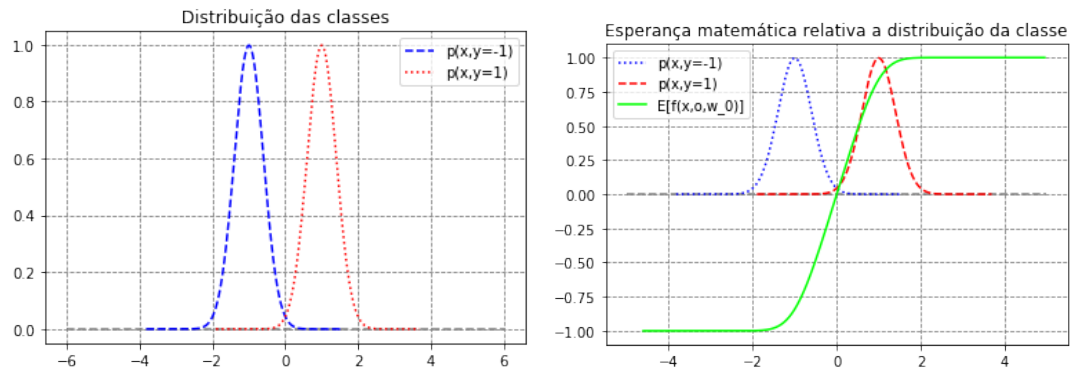
para gerar os exemplos de dados foram médias $\mu_- = -1$ e $\mu_+ = 1$. E variâncias $\sigma_- = \sigma_+ = 0.5$. Como mencionado anteriormente, o principal objetivo é calcular a esperança matemática da saída da função $f(x, o, w_0)$ segundo uma distribuição de probabilidade. Utilizando a distribuição gaussiana na expressão (??) a esperança matemática da saída de $f(\tilde{x}, o, w_0)$ será obtida como a seguir.

$$2\gamma_-(\tilde{x}) + 2\gamma_+(\tilde{x}) \tag{3.9}$$

Onde $\mathbb{E}[f(\tilde{x}, o, w_0)]$ seria a avaliação da integração (3.8) utilizando a equação resultante da integração da função densidade gaussiana. Onde $\gamma_{\pm}(\tilde{x})$ trata-se da expressão:

$$\gamma_{\pm}(\tilde{x}) = (\tilde{x} - \mu_{\pm}) \operatorname{erf} \left(\frac{(\tilde{x} - \mu_{\pm})}{\sqrt{2}\sigma_{\pm}} \right) + \sqrt{\frac{2}{\pi}} e^{-\left(\frac{(\tilde{x} - \mu_{\pm})}{\sqrt{2}\sigma_{\pm}}\right)^2} \tag{3.10}$$

Onde erf é a função erro, uma série aproximada obtida a partir da integral da gaussiana (Schuld e Petruccione 2018). Todas as réplicas das análises numéricas em (Schuld e Petruccione 2018) foram feitas utilizando a linguagem de programação *python*. Para se efetuar a réplica do caso unidimensional foi utilizada a biblioteca *numpy* (Oliphant 2006). Durante a replicação, foi observado que para que o valor da esperança matemática $\mathbb{E}[f(\tilde{x}, o, w_0)]$ permanecesse no intervalo $[-1, 1]$, o qual é o intervalo das classes, é necessário dividir cada termo do somatório (3.9) por 2 ao invés de multiplicá-los por 2 como os autores haviam escrito. E para que os exemplos de dados utilizados na replicação estivessem no intervalo de $(0, 1]$ fora necessário utilizar variâncias $\sigma_- = \sigma_+ = 0.4$ e não 0.5 como os autores havia descrito. Porém os valores para os parâmetros μ_- e μ_+ permaneceram inalterados.



(a) Distribuição dos exemplos de dados computados para o caso unidimensional (b) Esperança matemática para o caso do classificador avaliar uma saída nunca antes vista.

Figura 1 – Esperança matemática para o caso do classificador avaliar uma saída nunca antes vista. Cujos os dados estão na distribuição normal.

Ao se computar a esperança matemática da função $f(\tilde{x}, o, w_0)$, deseja-se computar para qual região de decisão tende a saída de cada exemplo de dados \tilde{x} ao se utilizar f . Ao se observar a Fig. 1.(a) Nota-se que a fronteira de decisão encontra-se no centro, entre a intersecção das distribuições das duas classes. Porém, ao se alterar as variâncias das distribuições de cada classe observa-se uma tendência do classificador em classificar exemplos de dados para uma das classes $\{-1, 1\}$. Servindo como uma evidência da dependência do desempenho de f à maneira como os dados estão distribuídos, como é mostrado na figura 2 .

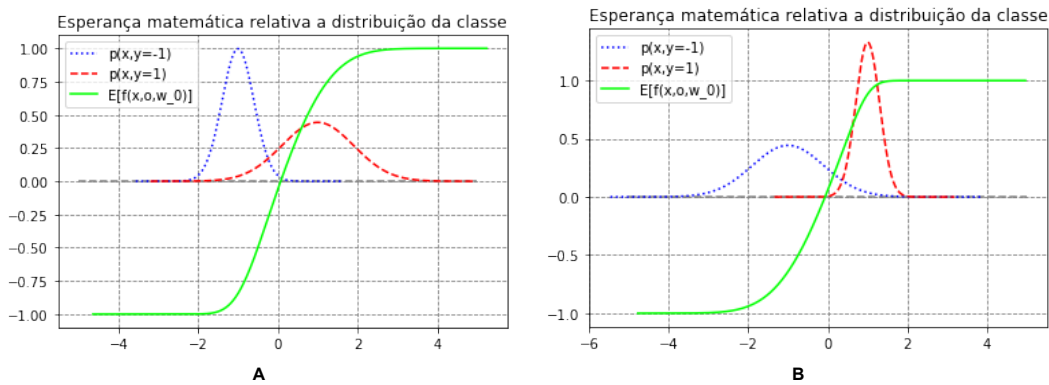


Figura 2 – **A:** $\mathbb{E}[f(\tilde{x}, o, w_0)]$ com variâncias $\sigma_- = 0.4$ e $\sigma_+ = 0.9$. **B:** $\mathbb{E}[f(\tilde{x}, o, w_0)]$ com variâncias $\sigma_- = 0.9$ e $\sigma_+ = 0.4$

A segunda parte da replicação tratou-se da aplicação de um comitê de classificadores em um conjunto de dados bidimensional básico. O principal objetivo é demonstrar a configuração das regiões de decisão em um espaço bidimensional, utilizando modelos perceptron simples, com função de ativação $sign(x)$. Onde saída do comitê é computada como na equação (3.1), onde a saída dos modelos são ponderadas pela acurácia dos mesmos. Para gerar os exemplos de dados para replicação do caso reduzido foi utilizada a função blob (*blob function*) da biblioteca *python scikit-learn* (Pedregosa et al. 2011) com os parâmetros utilizados como entrada da função para exemplos de dados gerados pertencentes a classe -1 foram $\mu_- = [-1, 1]$. E para exemplos de dados pertencentes à classe 1 foram $\mu_+ = [1, -1]$. A computação das regiões de decisão para o caso multidimensional foi feita a partir do uso do procedimento *meshgrid*. Com a resolução utilizada para computar as regiões de decisão foi de 0.5 . Foram utilizados 8000 modelos para se efetuar a classificação, com pesos e *bias* no intervalo $[-1, 1]$.

Para conjunto de dados pequenos, como na figura 3. Os comitê com os modelos utilizados separam bem os dados. A fronteira encontra-se próxima do centro onde ambos os conjuntos de dados se intercalam. Porém ao se mudar o desvio padrão de

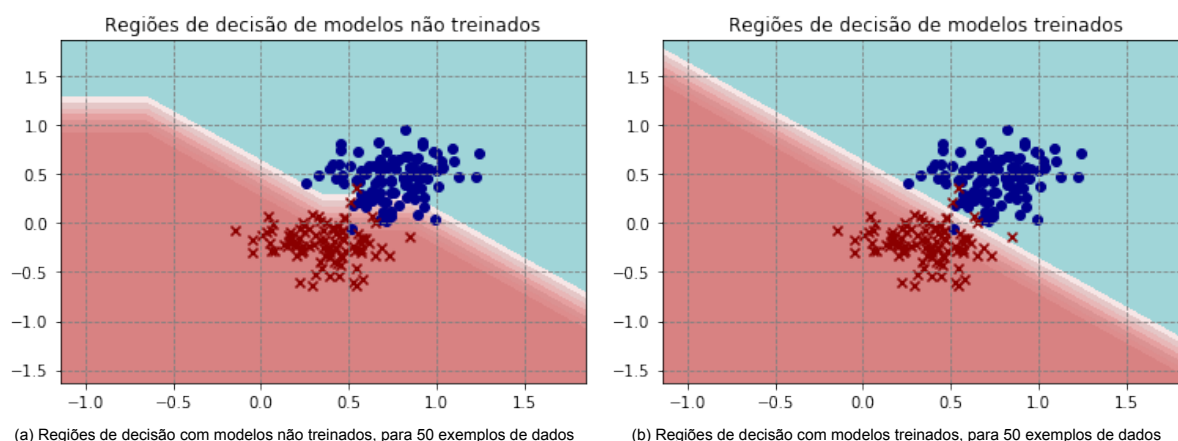


Figura 3 – Regiões de decisão para modelos treinados e não treinados utilizando os valores descritos de μ_{\pm} intervalos de pesos e número de modelos

uma das classes, a diferença entre o uso de modelos treinados e não treinados torna-se mais visível. Como na figura 4.(b) a fronteira de decisão encontra-se mais ajustada ao centro que separa os dois conjuntos do que a fronteira de decisão apresentada em 4.(a).

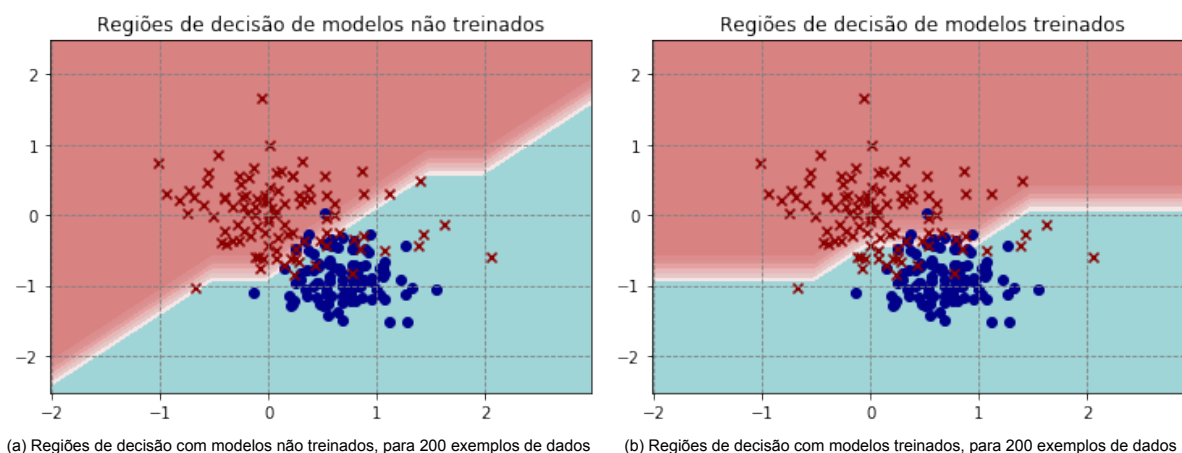


Figura 4 – Regiões de decisão para modelos treinados não treinados utilizando os valores descritos de μ_{\pm} intervalos de pesos e número de modelos

Porém estas foram somente replicações de exemplos simples apresentados em (Schuld e Petruccione 2018). Na seção 4 são apresentados resultados de um comitê de classificadores utilizando bases de dados reais.

3.2 Comitê quântico de classificadores

Esta descrição do funcionamento de comitês classificadores quânticos assume a existência de processadores quânticos com diversos qubits. Para a construção dos estados e processamento das informações foi considerado o fato das informações

serem inseridas nos estados utilizando *qubit-encoding* onde tanto os dados \tilde{x} como os classificadores θ ficam armazenados em registradores quânticos diferentes.

Desse modo o sistema quântico com *qubit-encoding* faria o uso de cinco registradores quânticos diferentes. Onde o primeiro registrador quântico $|x\rangle$ seria o registrador dos dados. O segundo registrador quântico $|\theta\rangle$ seria o registrador que codifica os parâmetros do classificador, ou como descrito no início da seção 3 o próprio classificador. O terceiro registrador quântico seria o registrador de um único qubit para o caso de classificação binária, ou mais de um caso contrário. Sendo o mesmo dedicado para o armazenamento da saída $|f(x, \theta)\rangle$, ou $|\hat{y}_\theta\rangle$. O quarto registrador seria dedicado ao armazenamento da saída desejada $|y\rangle$. E o quinto classificador seria um registrador auxiliar, também de um único qubit inicializado em super posição. Seja E o conjunto de classificadores utilizados no comitê, ou seja o conjunto de membros do comitê. O estado inicial para tal configuração do sistema ficaria como na equação (3.11).

$$|\psi\rangle = |x\rangle \frac{1}{\sqrt{|E|}} \sum_{\theta \in E} |\theta\rangle |0\rangle |y\rangle \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right) \quad (3.11)$$

Onde $|E|$ trata-se do tamanho do comitê. Os valores θ podem ser armazenados em superposição utilizando métodos já conhecidos de armazenamentos que foram utilizados em memórias associativas quânticas (Trugenberger 2001, Trugenberger 2002, Ventura e Martinez 1999). O mesmo pode ser dito para os valores de \tilde{x} , caso deseje-se armazenar todos os exemplos de dados em super posição. Por simplicidade, será considerado um armazenamento de \tilde{x} por vez. Seja \mathcal{F} o operador quântico que implementa a função $f(\tilde{x}, \theta)$. O operador \mathcal{F} seria aplicado no sistema utilizando os registradores $|x\rangle$ e $|\theta\rangle$ como entrada sendo a saída computada armazenada no terceiro registrador quântico.

$$\mathcal{F}|x\rangle \frac{1}{\sqrt{|E|}} \sum_{\theta \in E} |\theta\rangle |0\rangle |y\rangle \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right) \mapsto |x\rangle \frac{1}{\sqrt{|E|}} \sum_{\theta \in E} |\theta\rangle |\hat{y}_\theta\rangle |y\rangle \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right) \quad (3.12)$$

A próxima etapa é um ponto chave na utilização de comitês quânticos, pois visa possibilitar o sistema quântico a computar a saída segundo a equação (3.1). Para isso pode-se aplicar rotações como a R_y no último registrador quântico de modo que as amplitudes de probabilidade fiquem maiores próximas do estado $|0\rangle$ se $|\hat{y}_\theta\rangle = |y\rangle$. E amplitudes de probabilidade maiores para o estado $|1\rangle$ caso contrário. Dessa maneira, a acurácia dos classificadores estaria codificada na amplitude de probabilidade, e emaranhada com o estado do registrador quântico $|\theta\rangle$. Após processada toda a base de dados de treinamento, o estado do resultaria na equação (3.13). Como todos os dados da base de treinamento foram processados os registradores $|x\rangle$, $|\hat{y}_\theta\rangle$ e $|y\rangle$ podem ser

desconsiderados.

$$|\psi\rangle = \frac{1}{\sqrt{|E|}} \sum_{\theta \in E} |\theta\rangle (\sqrt{a_\theta}|0\rangle + \sqrt{1-a_\theta}|1\rangle) \quad (3.13)$$

Em seguida, uma medição deve ser feita no último registrador quântico. Onde, ao ser medido o mesmo se encontrar no estado $|1\rangle$ o processo deve ser retomado. Caso contrário, o sistema quântico construiu um comitê de classificadores quântico, cujo o número de membros do comitê pode ser exponencialmente grande, segundo o número de qubits disponíveis para o registrador $|\theta\rangle$. Seja \mathcal{X} um fator de normalização resultante da medição do último registrador quântico.

$$\sum_{\theta \in E} \frac{\sqrt{a_\theta}}{\sqrt{\mathcal{X}|E|}} |\theta\rangle \quad (3.14)$$

Onde o estado na equação 3.14 representa a superposição de todos os classificadores utilizados no comitê ($|\theta\rangle$), sendo o grau de influência de cada classificador codificado nas amplitudes de probabilidades $\sum_{\theta \in E} \frac{\sqrt{a_\theta}}{\sqrt{\mathcal{X}|E|}}$. Seja \tilde{x} uma entrada nunca vista. Uma vez obtido comitê de classificadores, o processo de classificar \tilde{x} necessitaria somente de dois registradores a mais, adicionados à equação (3.14). Sendo o primeiro destinado a armazenar os dados, e o segundo destinado a armazenar a saída dos classificadores. Cujas aplicação do operador \mathcal{F} é necessária para computar f segundo todos os classificadores contidos em E .

$$\mathcal{F}|\tilde{x}\rangle \sum_{\theta \in E} \frac{\sqrt{a_\theta}}{\sqrt{\mathcal{X}|E|}} |\theta\rangle |0\rangle \mapsto |\tilde{x}\rangle \sum_{\theta \in E} \frac{\sqrt{a_\theta}}{\sqrt{\mathcal{X}|E|}} |\theta\rangle |\hat{y}_\theta\rangle \quad (3.15)$$

Para se obter a classificação do comitê é necessário efetuar a medição no último qubit. Considerando o caso de classificação binária. Seja k o número de classificadores que retornaram a saída $|\hat{y}_\theta\rangle = |0\rangle$ e l o número de classificadores que retornaram a saída $|\hat{y}_\theta\rangle = |1\rangle$. As probabilidades de se obter os estados $|0\rangle$ ou $|1\rangle$ seriam segundo a equação (3.16). Uma desvantagem de um sistema quântico em superposição como na equação (3.15) é o fato de que ao se medir $|\hat{y}_\theta\rangle$ a superposição entraria em colapso. Logo, caso fosse desejado reutilizar o comitê, o passo a passo da construção do mesmo teria que ser refeito.

$$p(|\hat{y}_\theta\rangle = |0\rangle) = \sum_{\theta \in E} \frac{a_\theta}{\mathcal{X}|E|} \quad (3.16)$$

$$p(|\hat{y}_\theta\rangle = |1\rangle) = \sum_{\theta \in E} \frac{a_\theta}{\mathcal{X}|E|}$$

4 Experimentos e resultados

O uso de classificadores não treinados em um comitê, como uma forma de classificação livre de otimização de parâmetros é dependente de como os dados estão distribuídos. Apesar de alguns exemplos terem sido apresentados na seção 3, os mesmos se tratavam de casos reduzidos de análises numéricas feitas em (Schuld e Petruccione 2018). Esta seção mostra resultados de experimentos simulando o uso de comitê de classificadores quântico utilizando bases de dados *benchmark*.

Para fazer com que os modelos generalizem bem o problema de classificação em questão, foi considerado neste trabalho a adição de uma etapa de treinamento para os classificadores. Os classificadores utilizados neste trabalho foram redes *Multi Layer Perceptron* (MLP). Todos com 20 camadas escondidas e 200 neurônios em cada camada. Todos os modelos MLP foram implementados utilizando a biblioteca *pytorch* (Paszke et al. 2017), na linguagem de programação *python*. As bases de dados utilizadas para execução dos experimentos foram as disponíveis pela biblioteca *scikit learn* (Pedregosa et al. 2011). São elas *iris*, *wine* e *breast-cancer*.

O método de validação cruzada utilizado durante os experimentos foi o *hold-out*. Sendo os conjuntos de dados divididos em conjunto de treino com 70% dos dados e conjunto de validação com 30%. Cujos valores de acurácias correspondentes a cada subconjunto serão referidos como acurácias de treino e de validação respectivamente.

As simulações foram feitas computando as probabilidades dos estados, durante a construção do sistema e também após a medição dos *qubits* de resposta. Ao se efetuar os experimentos, o principal objetivo tratava-se de obter a acurácia geral de validação do comitê.

Uma outra métrica adotada foi a probabilidade por exemplo de dados (PPED). A qual era computada segundo o conjunto de validação, para cada exemplo de dados, onde a PPED tratava-se da probabilidade do comitê de classificadores retornar a classe correta ou não para um determinado exemplo de dados. Sendo a PPED para o comitê retornar a classe correta chamada de *Hit probability per sample*, a qual neste trabalho será referida como PPED de *Hit*. E a PPED para o comitê retornar a classe incorreta chamada de *Miss probability per sample*, a qual neste trabalho será referida como PPED de *Miss*. As saídas de um sistema quântico são obtidas de forma probabilística. Portanto, esta métrica foi adicionada para que fosse possível saber se ao medir o qubit $|\hat{y}\rangle$ presente no uso do comitê quântico, se o mesmo retornaria a classe correta com uma probabilidade alta. Ou se o mesmo retornaria uma classe incorreta com probabilidade alta.

Uma vez computadas as PPEDs, foi utilizada uma probabilidade limiar para que fosse possível computar a acurácia geral do comitê. Seja x_q um exemplo de dados qualquer do conjunto de validação, dada uma probabilidade limiar p_l , a classificação seria considerada um acerto do comitê se a probabilidade de *Hit* do exemplo x_q fosse maior ou igual a p_l . É a partir das comparações entre a probabilidade de *Hit* e probabilidade limiar que a *acurácia geral* do comitê de classificadores foi computada.

Durante as execuções o número de membros no comitê foi variando de acordo com o conjunto $|E| \in \{100, 200, 300, 400, 500\}$. As execuções foram divididas em duas etapas. Onde para cada base de dados utilizada e para cada $|E|$ fora executada a primeira etapa, onde consistia de membros do comitê não treinados. E a segunda etapa, a qual consistia de membros do comitê treinados. Primeiramente será abordado o uso de comitê quântico com classificadores não treinados. E em seguida serão apresentados os resultados do uso de comitê quântico com classificadores treinados. O algoritmo 1 apresenta uma versão alto nível do funcionamento das simulações.

Algoritmo 1: Pseudo código demonstrando o funcionamento das simulações

```

1 para Cada base de dados  $B \in \mathcal{B}$  faça
2   Criar o comitê de classificadores  $E = \{\theta_0, \dots, \theta_{|E|}\}$ ;
3   Dividir  $B$  em conjunto de treino ( $B_t$ ) e validação ( $B_v$ );
4   Inicializar o array  $PPED_{hit} \leftarrow \emptyset$ ;
5   Inicializar o array  $PPED_{miss} \leftarrow \emptyset$ ;
6   se Usar etapa de treinamento então
7     para Cada classificador  $\theta_i \in \{\theta_0, \dots, \theta_{|E|}\}$  faça
8       Treinar  $\theta_i$ ;
9       Calcular a acurácia  $a_\theta$ ;
10    fim
11  fim
12  Criar o array de amplitudes probabilidades  $\left[ \sqrt{\frac{a_{\theta_0}}{\chi^{|E|}}}, \dots, \sqrt{\frac{a_{\theta_{|E|}}}{\chi^{|E|}}} \right]$ ;
13  Processar  $B_v$  com dos classificadores  $\{\theta_0, \dots, \theta_{|E|}\}$ ;
14  para Cada exemplo de dados  $(x, y) \in B_v$  faça
15    Calcular a probabilidade  $p_{hit} = \sum_{\theta \in E} \sqrt{\frac{a_\theta}{\chi^{|E|}}}$  para cada classificador
       $\theta \in E$  que acertar  $y$ ;
16    Calcular a probabilidade  $p_{miss} = \sum_{\theta \in E} \sqrt{\frac{a_\theta}{\chi^{|E|}}}$  para cada classificador
       $\theta \in E$  que errar  $y$ ;
17    se  $p_{hit} \geq p_l$  então
18      Incrementar a acurácia geral do comitê;
19    fim
20    Salvar  $p_{hit}$  no array  $PPED_{hit}$ ;
21    Salvar  $p_{miss}$  no array  $PPED_{miss}$ ;
22  fim
23  Calcular a PPED média de Hit usando os dados do array  $PPED_{hit}$ 
24  Calcular a PPED média de Miss usando os dados do array  $PPED_{miss}$ 
25 fim

```

4.1 Comitê de classificadores não treinados quântico

Para o caso de um comitê com 100 classificadores não treinados, os valores de acurácia geral foi extremamente baixos para a base de dados *breast-cancer* e nulos para as bases de dados *iris* e *wine*. Os valores na tabela 1 representam valores de acurácia geral em relação a probabilidade limiar de 0.7. O limiar foi definido após uma sequência de testes envolvendo tentativa e erro, de modo a se alcançar um limiar alto suficiente de modo que um comitê que efetuasse rotulação aleatória não influenciasse no resultado e baixo o suficiente de modo que não fosse proibitivo para comitês que retornassem PPEDs de *Hit* razoavelmente altas (por exemplo, próximas porém não iguais a 90%) de serem computados. A acurácia média dos classificadores neste tipo de comitê foi de aproximadamente 0.29 com desvio padrão de aproximadamente 0.18.

Conjunto de dados	p_l	acurácia geral
<i>iris</i>	0.7	0.0
<i>wine</i>		0.0
<i>breast-cancer</i>		0.0058

Tabela 1 – Resultados para um comitê com 100 classificadores não treinados

Na figura 5 são mostrados gráficos de barra, que indicam as PPED de *Mis* e *Hit* para os exemplos de dados utilizando um comitê com 100 classificadores. Onde o eixo x dos gráficos indicam os exemplos de dados indexados por valores inteiros, no intervalo de $i \in [0 - (n - 1)]$. E o eixo y indica a probabilidade de se obter a classificação correta ou incorreta para i -ésimo exemplo de dados. As médias das PPEDs para um comitê com 100 classificadores não treinados foram de 0.4 de *Hit*, aproximadamente. E aproximadamente 0.6 de PPED média de *Miss*, para o conjunto de dados *iris*, ambas com desvio padrão de aproximadamente 0.38.

Para o conjunto de dados *wine* a PPED média de *Hit* foi de aproximadamente 0.41. E PPED média de *Miss* de aproximadamente 0.58, com desvio padrão de 0.057 para os casos de PPEDs médias de *Miss* e *Hit*.

E para o conjunto de dados *breast-cancer* a PPED média de *Hit* foi de aproximadamente 0.57. E PPED média de *Miss* de aproximadamente 0.42, com desvio padrão de 0.066 para os casos de PPEDs médias de *Miss* e *Hit*.

As tabelas 2, 3, 4 e 5 mostram situações cujos valores de acurácia geral dos comitês foram nulos utilizando probabilidade limiar de 0.7.

Utilizando o comitê com 200 classificadores não treinados como na tabela 2. Para o conjunto de dados *iris* a PPED média de *Hit* foi de aproximadamente 0.40. E PPED média de *Miss* de aproximadamente 0.59, com desvio padrão de 0.033 para os casos de PPEDs médias de *Miss* e *Hit*.

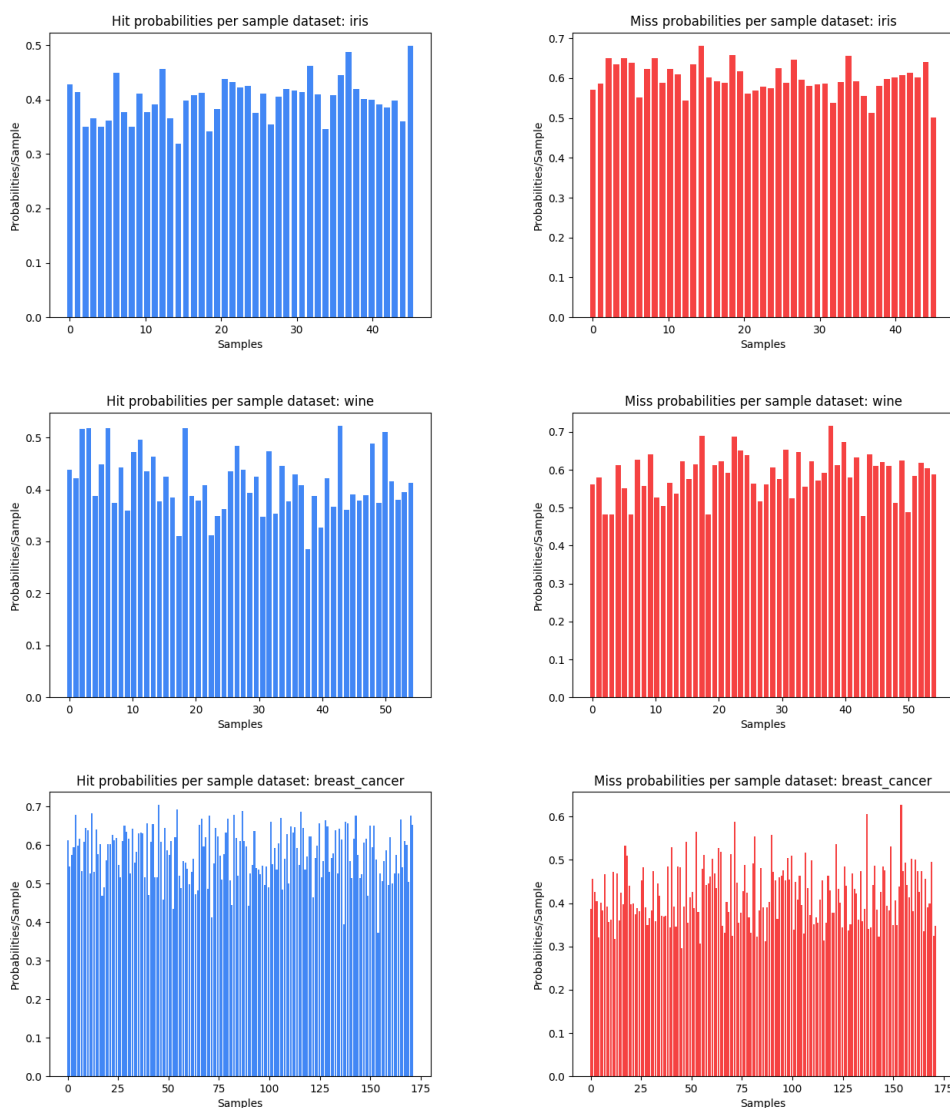


Figura 5 – PPEDs para as bases de dados *iris*, *wine* e *breast-cancer* utilizando o comitê quântico com 100 classificadores não treinados

Conjunto de dados	p_l	acurácia geral
<i>iris</i>		0.0
<i>wine</i>	0.7	0.0
<i>breast-cancer</i>		0.0

Tabela 2 – Resultados para um comitê com 200 classificadores não treinados

O conjunto de dados *wine* com a PPED média de *Hit* foi de aproximadamente 0.39. E PPED média de *Miss* de aproximadamente 0.6, com desvio padrão de 0.033 para os casos de PPEDs médias de *Miss* e *Hit*.

E o conjunto de dados *breast-cancer* com a PPED média de *Hit* foi de aproximadamente 0.56. E PPED média de *Miss* de aproximadamente 0.43, com desvio padrão de 0.06 para os casos de PPEDs médias de *Miss* e *Hit*.

Conjunto de dados	p_i	acurácia geral
<i>iris</i>		0.0
<i>wine</i>	0.7	0.0
<i>breast-cancer</i>		0.0

Tabela 3 – Resultados para um comitê com 300 classificadores não treinados

Utilizando o comitê com 300 classificadores não treinados como na tabela 3. O conjunto de dados *iris* com a PPED média de *Hit* foi de aproximadamente 0.41. E PPED média de *Miss* de aproximadamente 0.58, com desvio padrão de 0.035 para os casos de PPEDs médias de *Miss* e *Hit*.

O conjunto de dados *wine* com a PPED média de *Hit* foi de aproximadamente 0.39. E PPED média de *Miss* de aproximadamente 0.6, com desvio padrão de 0.029 para os casos de PPEDs médias de *Miss* e *Hit*.

E o conjunto de dados *breast-cancer* com a PPED média de *Hit* foi de aproximadamente 0.56. E PPED média de *Miss* de aproximadamente 0.43, com desvio padrão de 0.044 para os casos de PPEDs médias de *Miss* e *Hit*.

Conjunto de dados	p_i	acurácia geral
<i>iris</i>		0.0
<i>wine</i>	0.7	0.0
<i>breast-cancer</i>		0.0

Tabela 4 – Resultados para um comitê com 400 classificadores não treinados

Utilizando o comitê com 400 classificadores não treinados como na tabela 4. Para o conjunto de dados *iris* a PPED média de *Hit* foi de aproximadamente 0.43. E PPED média de *Miss* de aproximadamente 0.55, com desvio padrão de 0.042 para os casos de PPEDs médias de *Miss* e *Hit*.

O conjunto de dados *wine* com a PPED média de *Hit* foi de aproximadamente 0.38. E PPED média de *Miss* de aproximadamente 0.61, com desvio padrão de 0.035 para os casos de PPEDs médias de *Miss* e *Hit*.

E o conjunto de dados *breast-cancer* com a PPED média de *Hit* foi de aproximadamente 0.56. E PPED média de *Miss* de aproximadamente 0.43, com desvio padrão de 0.054 para os casos de PPEDs médias de *Miss* e *Hit*.

Conjunto de dados	p_i	acurácia geral
<i>iris</i>		0.0
<i>wine</i>	0.7	0.0
<i>breast-cancer</i>		0.0

Tabela 5 – Resultados para um comitê com 500 classificadores não treinados

Utilizando o comitê com 500 classificadores não treinados como na tabela 5.

O conjunto de dados *iris* com a PPED média de *Hit* foi de aproximadamente 0.41. E PPED média de *Miss* de aproximadamente 0.58, com desvio padrão de 0.064 para os casos de PPEDs médias de *Miss* e *Hit*.

O conjunto de dados *wine* a PPED média de *Hit* foi de aproximadamente 0.38. E PPED média de *Miss* de aproximadamente 0.61, com desvio padrão de 0.032 para os casos de PPEDs médias de *Miss* e *Hit*.

E o conjunto de dados *breast-cancer* a PPED média de *Hit* foi de aproximadamente 0.56. E PPED média de *Miss* de aproximadamente 0.43, com desvio padrão de 0.054 para os casos de PPEDs médias de *Miss* e *Hit*.

4.2 Comitê de classificadores treinados quântico

Nesta etapa, os membros foram treinados utilizando diferentes épocas de treinamento, com o número de épocas de acordo com o conjunto $\{100, 250, 500\}$. Em todas as épocas de treinamento e para todos os valores em $|E|$, o otimizador de parâmetros utilizado foi o *Stochastic Gradient Descent* (SGD) (Bottou 2010), com valor de taxa de aprendizado igual a 0.1 e *momentum* igual a 0.2, e o tamanho de *batch* utilizado igual a 10. Os parâmetros utilizados no caso de comitê de classificadores quânticos treinados foram definidos após uma sequência de testes envolvendo tentativa e erro. As tabelas 6, 8 e 10 mostram as acurácias gerais dos comitês de classificadores com diversos tamanhos para as bases de dados *iris*, *wine* e *breast-cancer*.

<i>iris</i>			
tamanho do comitê	nº épocas de treinamento	p_l	acurácia geral
100	100	0.7	0.93
	250		0.93
	500		1.0
200	100		0.95
	250		0.97
	500		0.93
300	100		0.95
	250		0.93
	500		0.97
400	100		0.93
	250		1.0
	500		0.93
500	100		0.97
	250		0.91
	500		0.97

Tabela 6 – Resultados para um comitê de classificadores quântico de diversos tamanhos para a base de dados *iris*

PPEDs médias e desvios padrões com a base de dados *iris*

tamanho do comitê	nº épocas	PPED média <i>Hit</i>	desvio padrão <i>Hit</i>	PPED média <i>Miss</i>	desvio padrão <i>Miss</i>
100	100	0.93	0.249	0.06	0.249
	250	0.94	0.217	0.05	0.217
	500	1.0	6.6×10^{-16}	0.0	0.0
200	100	0.96	0.194	0.04	0.194
	250	0.96	0.153	0.032	0.153
	500	0.94	0.204	0.05	0.204
300	100	0.96	0.182	0.03	0.182
	250	0.93	0.241	0.06	0.241
	500	0.98	0.106	0.01	0.106
400	100	0.93	0.218	0.06	0.218
	250	0.99	1.11×10^{-15}	0.0	0.0
	500	0.94	0.224	0.05	0.224
500	100	0.97	0.146	0.02	0.146
	250	0.92	0.256	0.07	0.256
	500	0.98	0.117	0.1	0.117

Tabela 7 – Probabilidades por exemplo de dados média e desvio padrão para os casos de acerto (*Hit*) e de erro (*Miss*) do comitê de classificadores quântico. Para os comitês de classificadores quânticos de diversos tamanhos, com classificadores treinados com 100, 250 e 500 épocas de treinamento. Usando a base de dados *iris*

Na tabela 7 são apresentadas as PPEDs médias e desvios padrões dos comitês quânticos de classificadores de diversos tamanhos para a base de dados de *iris*. Pode-se observar que as PPEDs médias de *Hit* apresentadas na tabela são consideravelmente maiores do que a PPED de *Hit* média calculada para o caso de um comitê quântico de classificadores não treinados. A maior PPED média calculada no caso de um comitê de classificadores quântico com 400 classificadores, treinados com 250 épocas de treinamento. Na figura 6 são mostrados gráficos de barras relacionados a PPED de *Hit* e *Miss* utilizando um comitê quântico com 400 classificadores. Onde pode-se observar que um comitê de classificadores quântico com esta configuração, cujos classificadores foram treinados durante 250 épocas de treinamento, retornaria a classe correta para todos os exemplos de dados do conjunto de validação com probabilidade próxima de 1.0.

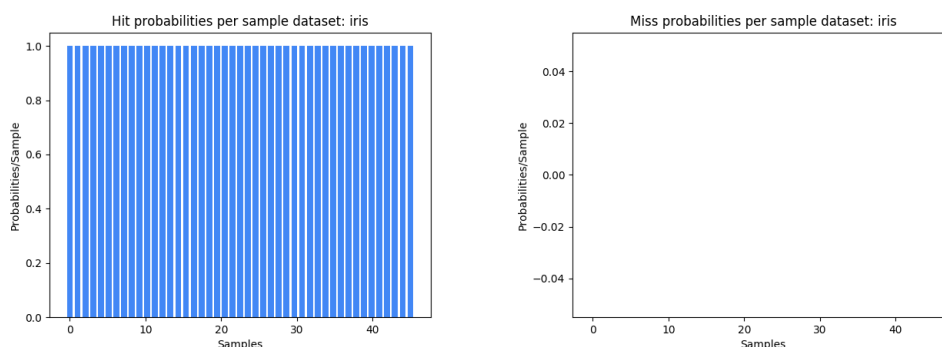


Figura 6 – PPEDs de *Hit* e *Miss* para as base de dados *iris* utilizando o comitê quântico com 400 classificadores treinados com 250 épocas de treinamento

wine

tamanho do comitê	nº épocas de treinamento	p_l	acurácia geral
100	100	0.7	1.0
	250		0.98
	500		1.0
200	100		0.98
	250		0.96
	500		0.96
300	100		0.98
	250		0.94
	500		0.96
400	100		0.96
	250		0.98
	500		1.0
500	100		0.98
	250		1.0
	500		0.98

Tabela 8 – Resultados para um comitê de classificadores quântico de diversos tamanhos para a base de dados *wine*

PPEDs médias e desvios padrões com a base de dados *wine*

tamanho do comitê	nº épocas	PPED média <i>Hit</i>	desvio padrão <i>Hit</i>	PPED média <i>Miss</i>	desvio padrão <i>Miss</i>
100	100	0.99	0.0013	0.0001	0.0013
	250	0.98	0.134	0.01	0.134
	500	0.99	0.004	0.0005	0.004
200	100	0.98	0.134	0.01	0.134
	250	0.96	0.182	0.03	0.182
	500	0.96	0.188	0.03	0.188
300	100	0.98	0.134	0.01	0.134
	250	0.96	0.165	0.03	0.165
	500	0.96	0.188	0.03	0.188
400	100	0.96	0.188	0.03	0.188
	250	0.97	0.137	0.02	0.137
	500	0.99	5.55×10^{-16}	0.0	0.0
500	100	0.98	0.134	0.01	0.134
	250	0.99	0.001	0.0002	0.001
	500	0.98	0.134	0.1	0.134

Tabela 9 – Probabilidades por exemplo de dados média e desvio padrão para os casos de acerto (*Hit*) e de erro (*Miss*) do comitê de classificadores quântico. Para os comitês de classificadores quânticos de diversos tamanhos, com classificadores treinados com 100, 250 e 500 épocas de treinamento. Usando a base de dados *wine*

Na tabela 9 são apresentadas as PPEDs médias dos comitês de classificadores de diversos tamanhos para a base de dados de *wine*. Assim como na tabela 7, as PPEDs médias de *Hit* na tabela 9 são consideravelmente maiores que as PPEDs de *Hit* médias apresentadas nos casos de comitês quânticos com classificadores não treinados. Sendo as maiores PPEDs médias calculadas para os casos de um comitês de classificadores quântico com 400 e 500 classificadores, treinados com 500 e 250 épocas de treinamento respectivamente. Na figura 6 são mostrados gráficos de barras

relacionados a PPED de *Hit* e *Miss* utilizando comitês quânticos com 400 e de 500 classificadores que apresentaram os melhores PPEDs médias de *Hit*. A partir da figura 7 pode-se observar que para o caso de um comitê quântico com 400 classificadores, o mesmo pode retornar a classe correta para todos os exemplos de dados com probabilidade próxima de 1.0. Apesar de haver alguns valores de probabilidades de *Miss*, o mesmo pode ser dito para o caso de um comitê quântico com 500 classificadores.

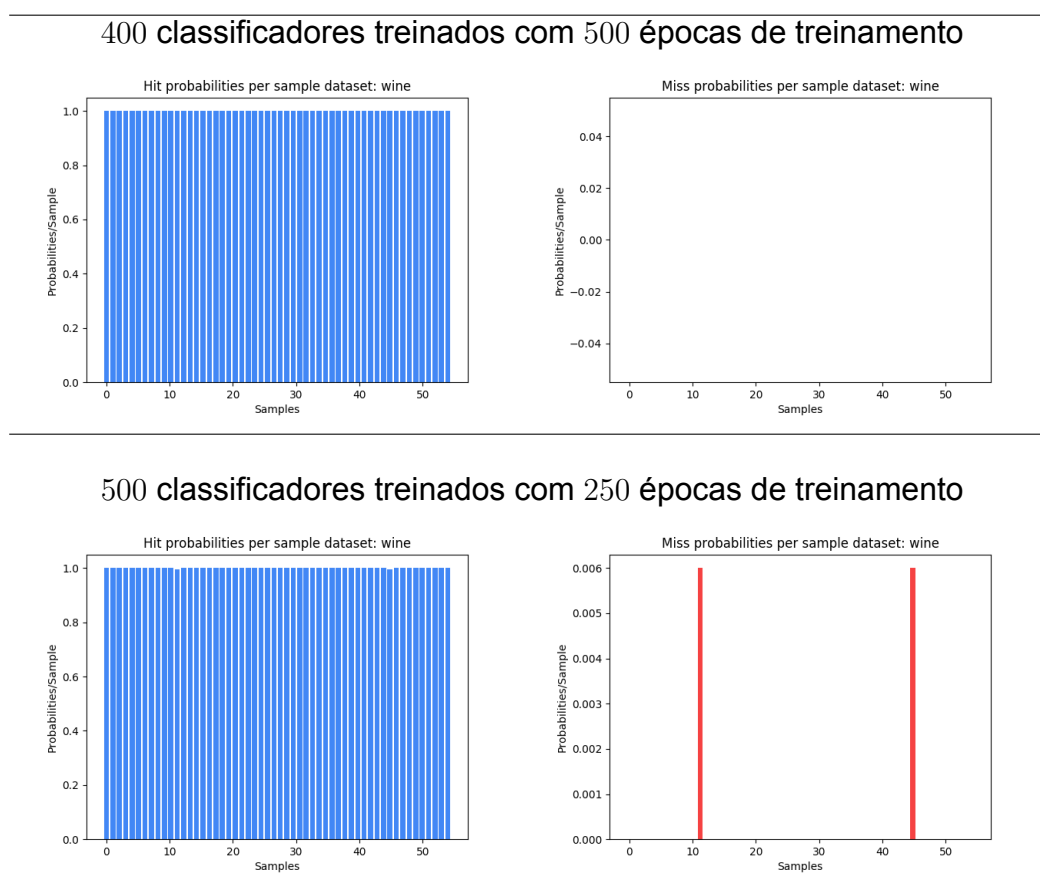


Figura 7 – PPEDs de *Hit* e *Miss* para as base de dados *wine* utilizando o comitê quântico com 400 e 500 classificadores treinados com 500 e 250 épocas de treinamento respectivamente

Na tabela 11 são apresentadas as PPEDs médias dos comitês de classificadores de diversos tamanhos para a base de dados de *breast-cancer*. Assim como na tabela 7 e 9 as PPEDs médias de *Hit* na tabela 11 também são consideravelmente maiores que as PPEDs médias apresentadas nos casos de comitês quânticos com classificadores não treinados. Sendo as maiores PPEDs de *Hit* médias calculadas para os casos de comitês de classificadores quânticos com 200 e 500 classificadores, para ambos os casos treinados com 500 épocas de treinamento. Na figura 8 são mostrados gráficos de barras relacionados a PPED de *Hit* e *Miss* utilizando comitês quânticos com 200 e de 500 classificadores que apresentaram os melhores PPEDs de *Hit* médias na tabela 11. Para o caso de um comitê quântico com 200 classificadores apresentados na figura 8 são consistentes com a acurácia geral de 0.97 apresentados na tabela

breast-cancer

tamanho do comitê	nº épocas de treinamento	p_l	acurácia geral
100	100	0.7	0.95
	250		0.97
	500		0.95
200	100		0.97
	250		0.96
	500		0.97
300	100		0.96
	250		0.96
	500		0.96
400	100		0.96
	250		0.95
	500		0.97
500	100		0.97
	250		0.97
	500		0.97

Tabela 10 – Resultados para um comitê de classificadores quântico de diversos tamanhos para a base de dados *breast-cancer*

PPDEs médias e desvios padrões com a base de dados *breast-cancer*

tamanho do comitê	nº épocas	PPDE média <i>Hit</i>	desvio padrão <i>Hit</i>	PPDE média <i>Miss</i>	desvio padrão <i>Miss</i>
100	100	0.95	0.193	0.04	0.193
	250	0.97	0.131	0.02	0.131
	500	0.96	0.168	0.03	0.168
200	100	0.97	0.148	0.02	0.148
	250	0.97	0.146	0.02	0.146
	500	0.98	0.119	0.01	0.119
300	100	0.97	0.153	0.02	0.153
	250	0.96	0.184	0.03	0.184
	500	0.96	0.17	0.03	0.17
400	100	0.96	0.178	0.03	0.178
	250	0.96	0.174	0.03	0.174
	500	0.97	0.136	0.02	0.136
500	100	0.97	0.151	0.02	0.151
	250	0.97	0.151	0.02	0.151
	500	0.98	0.133	0.01	0.133

Tabela 11 – Probabilidades por exemplo de dados média e desvio padrão para os casos de acerto (*Hit*) e de erro (*Miss*) do comitê de classificadores quântico. Para os comitês de classificadores quânticos de diversos tamanhos, com classificadores treinados com 100, 250 e 500 épocas de treinamento. Usando a base de dados *breast-cancer*

10 dado que existem exemplos de dados cuja probabilidade de *Miss* está acima de 0.7. O mesmo pode ser dito para o caso do comitê quântico com 500 classificadores apresentados na figura 8.

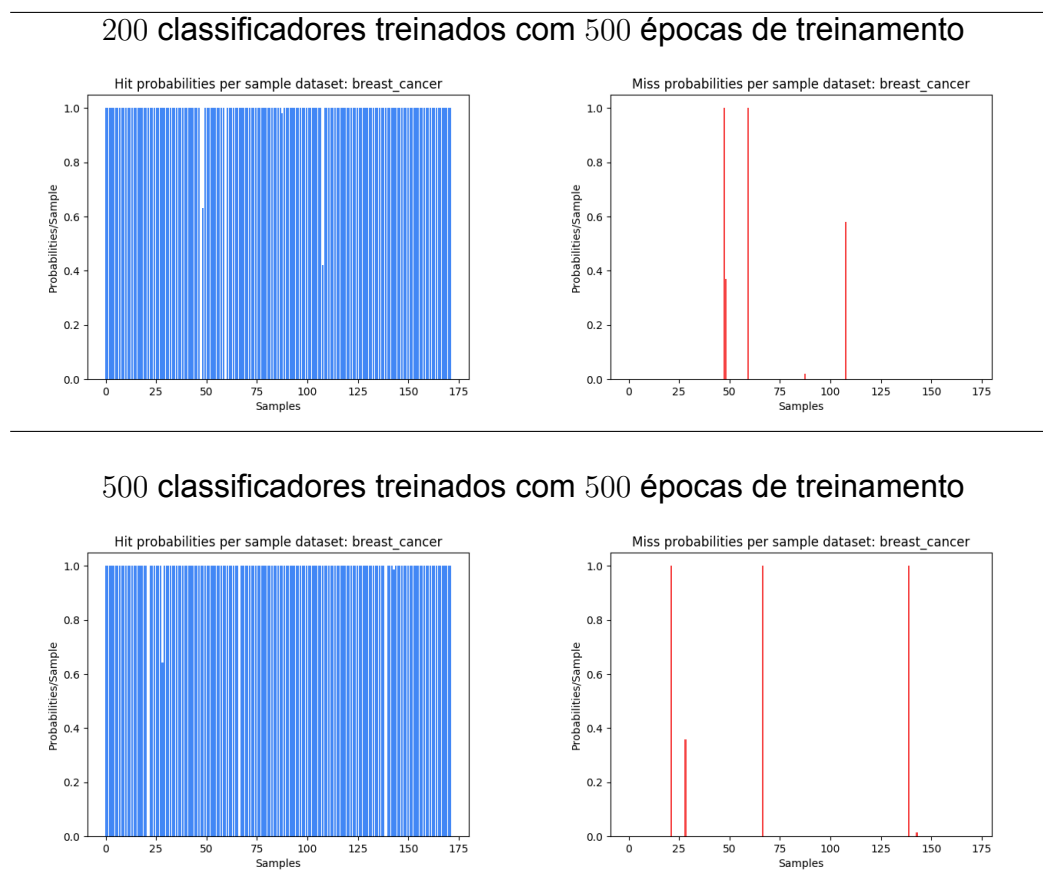


Figura 8 – PPEDs de *Hit* e *Miss* para as base de dados *wine* utilizando o comitê quântico com 200 e 500 classificadores, para ambos os casos treinados com 500 épocas de treinamento

4.3 Discussões

A partir dos resultados apresentados neste capítulo, pode-se concluir que a simples adição de uma etapa de treinamento aos classificadores aumentou de forma considerável o desempenho dos mesmos. Fazendo com que o uso de comitês de classificadores não treinados quânticos como uma forma de aprendizado independente de otimização de parâmetros uma escolha ruim para problemas de classificação. Devido a limitações hardware presentes em computadores quânticos atuais, foi escolhida a execução de experimentos simulados a partir do cálculo das probabilidades dos estados de um sistema quântico. Portanto, devido a essas mesmas limitações, não foi possível executar um comitê quântico de classificadores exponencialmente grande, como seria teoricamente possível em um processador quântico com um número de qubits suficientemente grande.

5 Conclusão

Este é um trabalho sobre o uso de comitês de classificadores quânticos e seu funcionamento ao se utilizar bases de dados *benchmark*. Apesar dos resultados serem oriundos de simulações a partir do cálculo da amplitude de probabilidade dos estados, as diferenças de desempenhos entre comitês quânticos de classificadores treinados e não treinados apresentadas nos resultados oferecem boas evidências sobre as vantagens da inclusão de uma etapa de treinamento na construção do comitê. Outros tipos de treinamentos também podem ser levados em consideração ao se utilizar comitês de classificadores treinados quânticos.

Os classificadores utilizados neste trabalho foram redes neurais artificiais e já existem trabalhos na literatura que investigam sobre codificação e treinamento de redes neurais em sistemas quânticos (Ricks e Ventura 2004, Silva, Ludermir e Oliveira 2016, Fawaz et al. 2019). Portanto, a ideia de se efetuar treinamento de redes neurais em um processador quântico não está muito distante do que seria teoricamente possível.

Neste trabalho foi considerada a inclusão de uma etapa de treinamento para um conjunto de redes neurais de arquitetura fixa, onde o termo arquitetura remete ao número de camadas escondidas e o número de neurônios em cada camada escondida. Em (Santos et al. 2018) é trabalhado não só com treinamento de redes neurais como também com a seleção de arquitetura.

Através de superposição e paralelismo quântico seria possível efetuar o treinamento de redes neurais com diferentes tipos de arquitetura utilizando todos os pesos possíveis que possam ser representados por um conjunto de *qubits* de tamanho n , onde as arquiteturas podem ser selecionadas a partir do desempenho.

Uma possibilidade de trabalho futuro seria executar experimentos incluindo a seleção de arquiteturas de redes neurais durante a etapa de treinamento, incluindo outras bases de dados, além das já utilizadas como *proben1* (Prechelt 1994) ou *imagenet* (Deng et al. 2009). Pelo fato dos resultados apresentados terem sido oriundos de simulações, os mesmos não levam em consideração ruído ou *de-coerência* de estados quânticos, situações em que sistemas quânticos verdadeiros sempre se encontrarão (Nielsen e Chuang 2000).

Já existem processadores quânticos que podem ser acessados na nuvem via interfaces de linguagens de programação, como é o caso do *framework Qiskit* (Aleksandrowicz et al. 2019) da empresa *IBM*. Porém os mesmos processadores tem uma quantidade de *qubits* muito pequena para que comitês de classificadores quânticos sejam reprodu-

zidos nos mesmos. Outra possibilidade de trabalho futuro seria a criação de um caso reduzido de comitê de classificadores quântico que possa ser executado em um dos processadores de pequena escala disponíveis.

Referências

- ALEKSANDROWICZ, G.; AL. et. *Qiskit: An Open-source Framework for Quantum Computing*. 2019. Citado na página 38.
- BIAMONTE, J. et al. Quantum machine learning. *Nature*, Nature Publishing Group, v. 549, n. 7671, p. 195, 2017. Citado na página 12.
- BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006. Citado 2 vezes nas páginas 11 e 12.
- BOTTOU, L. Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010*. [S.l.]: Springer, 2010. p. 177–186. Citado na página 32.
- DENG, J. et al. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*. [S.l.: s.n.], 2009. Citado na página 38.
- FACELI, K. et al. Inteligência artificial: Uma abordagem de aprendizado de máquina. 2011. Citado 2 vezes nas páginas 11 e 12.
- FAWAZ, A. et al. Training and meta-training binary neural networks with quantum computing. In: *ACM. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. [S.l.], 2019. p. 1674–1681. Citado 2 vezes nas páginas 14 e 38.
- FEYNMAN, R. P. Simulating physics with computers. *International journal of theoretical physics*, Springer, v. 21, n. 6, p. 467–488, 1982. Citado na página 12.
- GRANT, E. et al. Hierarchical quantum classifiers. *npj Quantum Information*, Nature Publishing Group, v. 4, n. 1, p. 65, 2018. Citado na página 12.
- HAYKIN, S. *Neural networks: principles and practice*. *Bookman*, 2001. Citado 2 vezes nas páginas 11 e 12.
- KEARNS, M.; VALIANT, L. Learning boolean formulae or finite automata is as hard as factoring. harvard university. *Center for Research in Computing Technology, Aiken Computation Laboratory*, 1988. Citado na página 13.
- MCMAHON, D. *Quantum computing explained*. [S.l.]: John Wiley & Sons, 2007. Citado 2 vezes nas páginas 16 e 17.
- MITCHELL, T. M. *Machine learning*. [S.l.]: McGraw hill, 1997. Citado na página 20.
- NIELSEN, M. A.; CHUANG, I. L. *Quantum computation and Quantum information*. [S.l.]: Cambridge University Press India, 2000. Citado 3 vezes nas páginas 11, 16 e 38.
- OLIPHANT, T. *Guide to NumPy*. [S.l.: s.n.], 2006. Citado na página 22.
- PASZKE, A. et al. Automatic differentiation in PyTorch. In: *NIPS Autodiff Workshop*. [S.l.: s.n.], 2017. Citado na página 27.

- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado 3 vezes nas páginas 14, 23 e 27.
- PRECHELT, L. Proben1: A set of neural network benchmark problems and benchmarking rules. Technical Report, 1994. Citado na página 38.
- RICKS, B.; VENTURA, D. Training a quantum neural network. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2004. p. 1019–1026. Citado na página 38.
- RUSSELL, S. J.; NORVIG, P. *Artificial intelligence: a modern approach*. [S.l.]: Malaysia; Pearson Education Limited,, 2016. Citado na página 11.
- SANTOS, P. dos et al. Quantum enhanced k-fold cross-validation. In: IEEE. *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.], 2018. p. 194–199. Citado 2 vezes nas páginas 14 e 38.
- SANTOS, P. G. D. et al. Quantum enhanced cross-validation for near-optimal neural networks architecture selection. *International Journal of Quantum Information*, World Scientific, v. 16, n. 08, p. 1840005, 2018. Citado na página 14.
- SCHAPIRE, R. E. The strength of weak learnability. *Machine learning*, Springer, v. 5, n. 2, p. 197–227, 1990. Citado na página 13.
- SCHULD, M.; FINGERHUTH, M.; PETRUCCIONE, F. Implementing a distance-based classifier with a quantum interference circuit. *arXiv preprint arXiv:1703.10793*, 2017. Citado na página 12.
- SCHULD, M.; PETRUCCIONE, F. Quantum ensembles of quantum classifiers. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 2772, 2018. Citado 8 vezes nas páginas 14, 15, 19, 20, 21, 22, 24 e 27.
- SILVA, A. J. da; LUDERMIR, T. B.; OLIVEIRA, W. R. de. Quantum perceptron over a field and neural network architecture selection in a quantum computer. *Neural Networks*, Elsevier, v. 76, p. 55–64, 2016. Citado na página 38.
- TOCCI, R. J.; WIDMER, N. S. *Sistemas Digitais: Princípios e Aplicações*. [S.l.]: Pearson-Prentice Hall, 11a edição.–John P. Uyemura, *Sistemas Digitais: uma ...*, 2002. Citado na página 18.
- TRUGENBERGER, C. A. Probabilistic quantum memories. *Physical Review Letters*, APS, v. 87, n. 6, p. 067901, 2001. Citado na página 25.
- TRUGENBERGER, C. A. Quantum pattern recognition. *Quantum Information Processing*, Springer, v. 1, n. 6, p. 471–493, 2002. Citado 2 vezes nas páginas 12 e 25.
- VENTURA, D. Artificial associative memory using quantum processes. In: *Proceedings of the International Conference on Computational Intelligence and Neuroscience*. [S.l.: s.n.], 1998. v. 2, p. 218–221. Citado na página 12.

VENTURA, D.; MARTINEZ, T. A quantum associative memory based on grover's algorithm. In: *Artificial Neural Nets and Genetic Algorithms*. [S.l.: s.n.], 1999. p. 22–27. Citado na página 25.

YANOFSKY, N. S.; MANNUCCI, M. A.; MANNUCCI, M. A. *Quantum computing for computer scientists*. [S.l.]: Cambridge University Press Cambridge, 2008. v. 20. Citado 2 vezes nas páginas 16 e 17.