Supplementary Material for "Quantum Perceptron Revisited: Computational-Statistical Tradeoffs"

Mathieu Roget[1,2] **Giuseppe Di Molfetta**[1] **Hachem Kadri**[1]

[1]Aix-Marseille University, CNRS, LIS, Marseille, France
[2]École Normale Superieure de Lyon, Lyon, France

1 PROOFS

In this appendix, we present the proofs of Theorems 4 and 5.

1.1 PROOF OF THEOREM 4

After proving a few useful lemma, we provide here the proof of the complexity of our HYBRID QUANTUM PERCEPTRON.

Lemma 1. *Let's define* $K = \left\lceil \frac{\ln(\epsilon/2)}{\ln(1-\sqrt{2}\gamma/\sqrt{\pi})} \right\rceil$, *then it holds that*

$$K \sim \sqrt{\frac{\pi}{2}} \frac{\ln(1/\epsilon)}{\gamma}.$$

Proof. Using a Taylor expansion for $\ln(1-x)$ in 0 we get

$$\begin{aligned}
\sqrt{\pi/2}\frac{\ln(1/\epsilon)}{K\gamma} &= \sqrt{\pi/2}\frac{\ln(1/\epsilon)\ln(1-\sqrt{2}\gamma/\sqrt{\pi})}{\gamma\ln(\epsilon/2)} \\
&= \sqrt{\pi/2}\frac{\ln(1/\epsilon)\left[-\sqrt{2}\gamma/\sqrt{\pi} + \underset{\gamma\to 0}{o}(\gamma)\right]}{\gamma\ln(\epsilon/2)} \\
&\underset{\gamma\to 0}{\to} \frac{\ln(1/\epsilon)}{\ln(1/\epsilon)+\ln(2)} \\
&\underset{\epsilon\to 0}{\to} 1.
\end{aligned}$$

Thus $K \sim \sqrt{\pi/2}\frac{\ln(1/\epsilon)}{\gamma}$. □

Lemma 2. *Let's define* $K2 = \left\lceil \log_{3/4}\left(1-\left(1-\frac{\epsilon}{2}\right)^{\frac{1}{K-1}}\right) \right\rceil$, *then it holds that*

$$K2 \sim \log_{3/4}(\epsilon\gamma).$$

Proof. Using a Taylor expansion for $\ln(1-\epsilon/2)$ and $\ln(1-\sqrt{\frac{2}{\pi}}\gamma)$ in 0 we get

$$(1-\epsilon/2)^{\frac{1}{K-1}} = \exp\left(\frac{\ln(1-\epsilon/2)\ln(1-\sqrt{\frac{2}{\pi}}\gamma)}{\ln(\epsilon/2)-\ln(1-\sqrt{\frac{2}{\pi}}\gamma)}\right) = \exp(-\alpha)$$

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

where

$$\alpha = \frac{1}{\sqrt{2\pi}} \frac{\epsilon\gamma}{\ln(\epsilon/2) - \ln(1-\sqrt{\frac{2}{\pi}}\gamma)} + o(\epsilon\gamma) \sim \frac{1}{\sqrt{2\pi}} \frac{\epsilon\gamma}{\ln(\epsilon/2) - \ln(1-\sqrt{\frac{2}{\pi}}\gamma)}.$$

Using $\ln(1-e^{-x}) \underset{x\to 0}{\sim} \ln(x)$, it holds that

$$K_2 = \log_{3/4}\left(1-e^{-\alpha}\right) \sim \log_{3/4}(\alpha) \sim \log_{3/4}(\epsilon\gamma).$$

□

Theorem 4. *Let S be a linearly separable sample of N points of margin γ. Algorithm* HYBRID QUANTUM PERCEPTRON *finds a perfect separator with probability at least $1-\epsilon$ and has a complexity of*

$$O\left(\frac{\sqrt{N}}{\gamma}\ln(1/\epsilon)\ln\left(\frac{1}{\gamma\epsilon}\right)\right).$$

Proof. The algorithm can fail because of two reasons. It is possible that none of the hyperplanes w_i, $i = 1, \ldots, K$, separate the classes and it is also possible that the quantum search gives a wrong result.

The exact value of K we take is $K = \left\lceil \frac{\ln(\epsilon/2)}{\ln(1-\sqrt{2}\gamma/\sqrt{\pi})} \right\rceil = O\left(\frac{\ln(1/\epsilon)}{\gamma}\right)$ because of lemma 2. The probability that a randomly drawn hyperplane separates the data is $\sqrt{2/\pi}\gamma$ (from Wiebe et al., 2016, Proof of theorem 2). Thus, the probability that at least one hyperplane separates the classes is

$$\mathbb{P}(\text{separating } w \text{ exists}) = 1 - \left(1-\sqrt{\frac{2}{\pi}}\gamma\right)^K \geq \left(1-\sqrt{\frac{2}{\pi}}\gamma\right)^{\frac{\ln(\epsilon/2)}{\ln(1-\sqrt{2}\gamma/\sqrt{\pi})}} = 1 - \frac{\epsilon}{2}.$$

Next we will assume that one of the K hyperplanes separates the classes. The algorithm will still return a wrong answer if it identifies a non-separating hyperplane as a separating one. The worst case is when the separating hyperplane is the K^{th} one. The probability that $K-1$ non-separating hyperplanes are all correctly identified is

$$\left(1-\frac{3}{4}^{K_2}\right)^{K-1} \geq 1-\frac{\epsilon}{2},$$

where

$$K_2 = \left\lceil \log_{3/4}\left(1-\left(1-\frac{\epsilon}{2}\right)^{\frac{1}{K-1}}\right)\right\rceil = O\left(\ln(1/(\gamma\epsilon))\right) \text{ (from lemma 2).}$$

The probability of failure is then bounded by

$$\mathbb{P}(\text{failure}) \leq \underbrace{\frac{\epsilon}{2}}_{\text{separating } w \text{ doesn't exist}} + \underbrace{\frac{\epsilon}{2}}_{\text{one non-separating hyperplane misidentified}} = \epsilon$$

and the complexity is

$$O\left(KK_2\sqrt{N}\right) = O\left(\frac{\sqrt{N}}{\gamma}\ln(1/\epsilon)\ln\left(\frac{1}{\gamma\epsilon}\right)\right)$$

which concludes the proof. □

1.2 PROOF OF THEOREM 5

For proving Theorem 5, the following definition and lemma are useful.

Definition 1. We define the Leave-one-out (LOO) error on a dataset S by

$$\hat{R}_{LOO}(S) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{h_{S-\{x_i\}}(x_i) \neq y_i\} \ , \tag{1}$$

where $h_{S-\{x_i\}}$ is the hypothesis returned by HYBRID QUANTUM PERCEPTRON on $S - \{x_i\}$, which is the same as S except that x_i has been deleted.

The lemma below shows the link between the expected risk and the Leave-one-out error.

Lemma 3 (Mohri et al., 2018, Lemma 5.3)**.** *For any* $N \geq 1$*,*

$$\mathop{\mathbb{E}}_{S\sim\mathcal{D}^N} [R(h_S)] = \mathop{\mathbb{E}}_{S'\sim\mathcal{D}^{N+1}} [\hat{R}_{LOO}(S')] \ .$$

Theorem 5. *Assume that the data is linearly separable. Let* h_S *be the hypothesis returned by the* HYBRID QUANTUM PERCEPTRON *algorithm after training over a sample* S *of size* N *drawn according to some distribution* $\mathcal{D}$*. Then, the expected error of* h_S *is bounded as follows:*

$$\mathbb{E}_{S\sim\mathcal{D}^N}\left(R(h_S)\right) \leq \sqrt{\frac{\pi}{2}} \frac{\log 1/\epsilon}{N+1} \mathbb{E}_{S\sim\mathcal{D}^{N+1}} \left(\frac{1}{\gamma_S}\right) \ .$$

Proof. The proof is based on computing an upper bound of the Leave-one-out error. Since the hyperplanes are drawn beforehand, they are the same for all instances $(S - \{x_i\})_i, \forall i = 1, \dots, N$. We also assume that there is at least one hyperplane that separates the training set S of size N (true with probability $1 - \epsilon$). If $N \leq K$ then the number of errors in $\hat{R}_{LOO}$ is naturally bounded by $N \leq K$ so it holds that $\hat{R}_{LOO} \leq K/N$. Thus we can restrict ourselves to the non trivial case where $K < N$.

We know that there is an hyperplane that separates the training set S correctly. Apart this hyperplane, noted w_K, the worst scenario is when the other ones all classify correctly all the data except one. Without loss of generality we consider that each w_k misclassifies only $x_k, \forall 1 \leq i < K$. So we will have one error for each of the $K - 1$ first predictions. Now, when HYBRID QUANTUM PERCEPTRON is trained on $S - \{x_i\}, \forall K \leq i \leq N$, the algorithm will choose the hyperplane w_K because it is the only one that correctly separates $S - \{x_i\}$ for $i = K, \dots, N$. Since w_K is the hyperplane returned by HYBRID QUANTUM PERCEPTRON on all the sample S, it will also correctly classify the points $x_i, \forall K \leq i \leq N$. Hence it holds that

$$\hat{R}_{LOO} \leq \frac{K}{N} \ .$$

Using Lemma 3 and $K \sim \sqrt{\frac{\pi}{2}} \frac{\ln(1/\epsilon)}{\gamma}$ (lemma 1), we obtain

$$\mathbb{E}_{S\sim\mathcal{D}^N}\left(R(h_S)\right) \leq \sqrt{\frac{\pi}{2}} \frac{\log 1/\epsilon}{N+1} \mathbb{E}_{S\sim\mathcal{D}^{N+1}} \left(\frac{1}{\gamma_S}\right) \ .$$

□

References

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2nd edition, 2018.

Nathan Wiebe, Ashish Kapoor, and Krysta M Svore. Quantum perceptron models. In *NIPS*, 2016.