More Midterm 2 Practice Problems

Problem 1. For each statement below, circle `True` if the statement is known to be true, `False` if the statement is known to be false, and `Open` if the statement is not known to be either true or false. Ensure that you pay careful attention to the formal definitions of asymptotic notation in your responses.

1. `True` `False` `Open` For learning problems that are linearly separable, the fastest possible algorithm for minimizing the 0-1 loss runs in sub-exponential time in the number of feature dimensions d.

2. `True` `False` `Open` For learning problems that are not linearly separable, the fastest possible algorithm for minimizing the 0-1 loss runs in sub-exponential time in the number of feature dimensions d.

3. `True` `False` `Open` Let $\mathcal{H}$ be the perceptron hypothesis class. Assume there exists some $h \in \mathcal{H}$ with $E_{\text{out}}(h) = 0$. Then after a finite number of iterations, the PLA is guaranteed to output a hypothesis g satisfying $E_{\text{in}}(g) = 0$.

4. `True` `False` `Open` Let $\mathcal{H}$ be the perceptron hypothesis class. Assume there exists some $h \in \mathcal{H}$ with $E_{\text{in}}(h) = 0$. Then after a finite number of iterations, the PLA is guaranteed to output a hypothesis g satisfying $E_{\text{out}}(g) = 0$.

5. `True` `False` `Open` Let $\mathcal{H}$ be the perceptron hypothesis class. Assume there exists some $h \in \mathcal{H}$ with $E_{\text{in}}(h) = 0$. Then the PLA will terminate.

6. `True` `False` `Open` Let $\mathcal{H}$ be the perceptron hypothesis class. Let $g_1 \in \mathcal{H}$ be the output of the PLA and $g_2 \in \mathcal{H}$ be the output of the pocket algorithm, and assume that both algorithms successfully terminate in finite time. The VC theory predicts that g_2 will have better generalization error than g_1 with high probability.

7. `True` `False` `Open` Let $\mathcal{H}$ be the perceptron hypothesis class and let $\mathcal{H}_\Phi$ be the perceptron hypothesis class with the 3rd degree polynomial feature map applied. If you attempt to use the PLA to select a hypothesis $g \in \mathcal{H}$ and the algorithm terminates, then the PLA is also guaranteed to terminate if you use it to select a hypothesis from $\mathcal{H}_\Phi$.

8. `True` `False` `Open` If your dataset is linearly separable, then the PLA is guaranteed to terminate.

9. `True` `False` `Open` If your dataset is not linearly separable, then the pocket algorithm is guaranteed to terminate.

10. `True` `False` `Open` Let $\mathcal{H}$ be the perceptron hypothesis class and let $g \in \mathcal{H}$ be the output of the pocket algorithm. Then the Hoeffding inequality can be used to bound the generalization error of g (i.e. the Hoeffding inequality can be used to bound $|E_{\text{in}}(g) - E_{\text{out}}(g)|$).

11. `True` `False` `Open` Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two hypothesis classes with $d_{\text{VC}}(H_1) > d_{\text{VC}}(H_2)$. Let $g_1 \in H_1$ and $g_2 \in H_2$. Hoeffding's inequality predicts that $|E_{\text{test}}(g_1) - E_{\text{out}}(g_1)|$ is less than $|E_{\text{test}}(g_2) - E_{\text{out}}(g_2)|$.

12. `True` `False` `Open` Let

$$\mathcal{H}_{\text{axis2}} = \Big\{ \mathbf{x} \mapsto \sigma \operatorname{sign}(x_i) : \sigma \in \{+1, -1\}, i \in [d] \Big\},$$

and

$$\mathcal{H}_{\text{L2-4}} = \Big\{ \mathbf{x} \mapsto [\![\|\mathbf{x}\|_2 \geq \alpha]\!] : \alpha \in \{1,2,3,4\} \Big\}.$$

Let $g_{\text{axis2}} \in \mathcal{H}_{\text{axis2}}$ and $g_{\text{L2-4}} \in \mathcal{H}_{\text{L2-4}}$ be the outputs of the TEA algorithm on their respective hypothesis classes. Then for learning problems with large d, the following inequality is guaranteed to hold: $E_{\text{in}}(g_{\text{axis2}}) \leq E_{\text{in}}(g_{\text{L2-4}})$.

13. `True` `False` `Open` Let

$$\mathcal{H}_{\text{axis2}} = \Big\{ \mathbf{x} \mapsto \sigma \operatorname{sign}(x_i) : \sigma \in \{+1, -1\}, i \in [d] \Big\},$$

and

$$\mathcal{H}_{\text{L2-4}} = \Big\{ \mathbf{x} \mapsto [\![\|\mathbf{x}\|_2 \geq \alpha]\!] : \alpha \in \{1,2,3,4\} \Big\}.$$

Let $g_{\text{axis2}} \in \mathcal{H}_{\text{axis2}}$ and $g_{\text{L2-4}} \in \mathcal{H}_{\text{L2-4}}$ be the outputs of the TEA algorithm on their respective hypothesis classes. Then for learning problems with large d, VC theory predicts that $g_{\text{L2-4}}$ will have better generalization accuracy than g_{axis2}.

14. `True` `False` `Open` The VC dimension of finite hypothesis classes can never be ∞.

15. `True` `False` `Open` The VC dimension of infinite hypothesis classes can never be ∞.

16. `True` `False` `Open` Let $\mathcal{H}$ be a hypothesis class. If there exists a hypothesis $h \in \mathcal{H}$ such that $E_{\text{out}}(h) = 0$, then the VC dimension of $\mathcal{H}$ must be finite.

17. `True` `False` `Open` Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two hypothesis classes satisfying $\mathcal{H}_1 \subset \mathcal{H}_2$. Then the approximation error of $\mathcal{H}_1$ is guaranteed to be less than or equal to the approximation error of $\mathcal{H}_2$.

18. True False Open For every learning problem, the approximation error is less than or equal to the in-sample error.

19. True False Open You have a learning problem with a high approximation error. VC theory predicts that increasing the number of data points will reduce the out-of-sample error.

20. True False Open You have a learning problem with a high approximation error. Applying the polynomial feature embedding with a high degree will decrease the approximation error.

21. True False Open You have a learning problem with a high approximation error. Applying the random feature embedding with a low output degree will decrease the approximation error.

22. True False Open You have trained a logistic regression model that has high generalization error. VC theory predicts that applying the PCA feature embedding with a low output dimension will reduce the generalization error.

23. True False Open Let $\mathcal{H}$ be the perceptron hypothesis class and let $\mathcal{H}_\Phi$ be the perceptron hypothesis class with the decision stump feature map. VC theory predicts that the generalization error of $\mathcal{H}_\Phi$ will be better than the generalization error of $\mathcal{H}$.

24. True False Open Let $\mathcal{H}$ be the perceptron hypothesis class and let $\mathcal{H}_\Phi$ be the perceptron hypothesis class with the decision stump feature map. Let $g \in \mathcal{H}$ and $g_\Phi \in \mathcal{H}_\Phi$ be the emperical risk minimizers trained on a very large dataset. VC theory predicts that $E_{\text{out}}(g) < E_{\text{out}}(g_\Phi)$.

25. True False Open The hinge loss is a surrogate loss function.

26. True False Open The trimmed hinge loss is convex.