More Midterm 3 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` If your model is overfitting, then a reasonable strategy for improving performance is to increase the size of your VC dimension in order to make E_{in} smaller.

2. `True` `False` `Open` If your model is underfitting, then a reasonable strategy for improving performance is to increase the size of your VC dimension in order to make E_{in} smaller.

3. `True` `False` `Open` If your model has low in-sample error E_{in} but high generalization error $|E_{\text{in}} - E_{\text{out}}|$, then a reasonable strategy for improving performance is to increase the number of data points N in the training set.

4. `True` `False` `Open` If your model is overfitting, then a reasonable strategy for improving performance is to increase the number of feature dimensions d in the training set.

5. `True` `False` `Open` You have trained a logistic regression model with L2 regularization. If your training set size N increases, then the optimal soft order constraint regularization hyperparameter C will also increase.

6. `True` `False` `Open` You have trained a logistic regression model with L1 regularization. If your training set size N increases, then the optimal augmented error regularization hyperparameter λ will also increase.

7. `True` `False` `Open` You have trained a logistic regression model with L1 regularization. If your number of feature dimensions d increases, then the optimal soft order constraint regularization hyperparameter C will also increase.

8. `True` `False` `Open` You have trained a logistic regression model with elastic net regularization. If your number of feature dimensions d increases, then the optimal augmented error regularization hyperparameter λ will also increase.

9. `True` `False` `Open` When training a logistic regression model, if you want the weight vector to be sparse, then you should prefer L1 regularization to L2 regularization.

10. `True` `False` `Open` When training a logistic regression model using L2 regularization, increasing the value of the augmented error regularization hyperparameter λ increases the VC dimension.

11. `True` `False` `Open` You have trained a logistic regression model with the PCA kernel and used a validation set to determine that the optimal output dimension k is 200. If your training set size N increases, then VC theory predicts that the optimal value for k will increase.

12. `True` `False` `Open` You have trained a MLP with the ReLU activation function and used a validation set to determine that the optimal number of layers is 5 and optimal width is 100. If your training set size N increases, and you keep the number of layers the same, then VC theory predicts that the optimal width of those layers will increase.

13. `True` `False` `Open` You have trained a boosted decision stump model and used a validation set to determine that the optimal number of base classifiers T is 1000. If instead of training a decision stump, you train a decision tree of depth 2, then the optimal number of base classifiers will increase.

14. `True` `False` `Open` The VC dimension of neural networks with the ReLU activation function is $\Theta(Ek\log(E))$, where k and E is as-defined in the notes.

15. `True` `False` `Open` The VC dimension of neural networks with the identity activation function is $O(dE)$.

16. `True` `False` `Open` Assume you are training a boosted decision stump model on a dataset with $N = 10^6$ and $d = 10^6$. Then in the limit as the number of base models T approaches infinity, the training error is guaranteed to approach 0 for all possible datasets.

17. `True` `False` `Open` Assume you are training an SVM with the random features kernel with output dimension d'. Then in the limit as d' approaches infinity, the training error is guaranteed to approach 0 for all possible datasets.

18. `True` `False` `Open` In vowpal wabbit, increasing the `--bit_precision` hyperparameter increases the model's VC dimension.

19. `True` `False` `Open` In vowpal wabbit, increasing the `--hash_seed` hyperparameter increases the model's VC dimension.

20. `True` `False` `Open` In vowpal wabbit, increasing the `--ngrams` hyperparameter increases the model's VC dimension.

21. `True` `False` `Open` In scikit-learn's `sklearn.tree.DecisionTreeClassifier` model, increasing the value of the hyperparamater `max_depth` increases the VC dimension.

22. `True` `False` `Open` You have trained a scikit-learn `sklearn.tree.DecisionTreeClassifier` model, but it is underfitting. VC theory predicts that increasing the value of `min_samples_split` will be more likely to improve performance than decreasing this value.

23. `True` `False` `Open` You are training a scikit-learn `sklearn.ensemble.AdaBoostClassifier` model with `base_estimator` set to `sklearn.tree.DecisionTreeClassifier`. In order to keep the VC dimension of your model constant, if you increase the value of `n_estimators` for the `AdaBoostClassifier`, then you should also increase the value of `max_depth` for the `DecisionTreeClassifier`.

24. `True` `False` `Open` You are training a scikit-learn `sklearn.ensemble.AdaBoostClassifier` model with `base_estimator` set to `sklearn.tree.DecisionTreeClassifier`. In order to keep the VC dimension of your model constant, if you increase the value of `min_samples_split` for the `DecisionTreeClassifier` then you should also increase the value of `n_estimators` for the `AdaBoostClassifier`.

25. `True` `False` `Open` You are using transfer learning to train the final layer of a deep neural network for your specific task. The `ResNet50` model has 50 hidden layers and the `ResNet18` model has only 18 hidden layers, and in both cases all layers have the same width. VC theory predicts that if you use the `ResNet50` model you will have a higher generalization error than if you use the `ResNet18` model.

26. `True` `False` `Open` You are using transfer learning to train the final layer of a deep neural network for your specific task. The `VGG13` model has 13 hidden layers and the `VGG19` model has 19 hidden layers, and in both cases all layers have the same width. VC theory predicts that if you use the `VGG13` model you will have a higher generalization error than if you use the `VGG19` model.