

## EECS545 Lecture 7 Quiz Solutions

1. Overfitting is characterized by:
  - (a) Low variance and low bias
  - (b) Low variance and high bias
  - (c) High variance and low bias
  - (d) High variance and high bias

**Solution:** (c). Overfitting is the case when there is a high variance and low bias.

2. Consider polynomial regression by optimizing the least-squares objective function with regularization:  $\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (\mathbf{w}^\top \phi(x^{(i)}) - y^{(i)})^2 + \lambda \|\mathbf{w}\|^2$  where  $x^{(i)}$  is scalar and  $\phi(x^{(i)})_j = (x^{(i)})^j$ . **Choose all options that apply:**
  - (a) It is not valid to use the  $L1$  norm in place of the  $L2$  norm in the regularized objective function.
  - (b) Increasing  $\lambda$  increases estimator bias and decreases estimator variance.
  - (c) Best practice indicates that we should keep tuning  $\lambda$  until we find a final value  $\lambda^*$  that minimizes error on the test set.
  - (d) As  $\lambda$  increases, the curve formed by polynomial regression will become flatter.

**Solution:** (b) and (d). (a) You can choose L1 if it is better suited for the problem you are trying to solve. (c) You must not use the test set to determine the value of a hyper-parameter.

3. (True/False) Training with more training data usually improves generalization. Suppose we have a training set containing  $N$  training cases. Suppose we duplicate the training cases  $L$  times each so that the training set now contains  $L \times N$  cases (but still only  $N$  unique cases). Will this improve the generalization performance of a logistic regression (in any case, assume that the regularization (e.g.,  $L_2$ ) hyperparameter will be carefully tuned on the validation data)?

**Solution:** False. Naively duplicating data does not help since the objective function will be increased by a constant factor. So, the optimal solution would remain the same as in the case of using original data.

4. (True/False) Suppose you are using logistic regression (trained with maximum likelihood), and the classifier is performing poorly (has an unacceptably high error) on the test set but performing well on the training set. In this case, adding more training examples will likely increase the training error. (For simplicity, we assume that both the training and testing examples were randomly sampled from an IID distribution.)

**Solution:** True. because it increases the difficulty of fitting (compared to the original setting)

5. (True/False) (Continued from Q4) Adding more training examples will likely increase the test error.

**Solution:** False. The model has seen more training data and may cover the data space more densely and thus generalize better.