



Ridge Regression

9 questions

1
point

1.

Which of the following is NOT a valid measure of overfitting?

- ☐ Sum of parameters ($w_1 + w_2 + \dots + w_n$)
- ☐ Sum of squares of parameters ($w_1^2 + w_2^2 + \dots + w_n^2$)
- ☐ Range of parameters, i.e., difference between maximum and minimum parameters
- ☐ Sum of absolute values of parameters ($|w_1| + |w_2| + \dots + |w_n|$)

1
point

2.

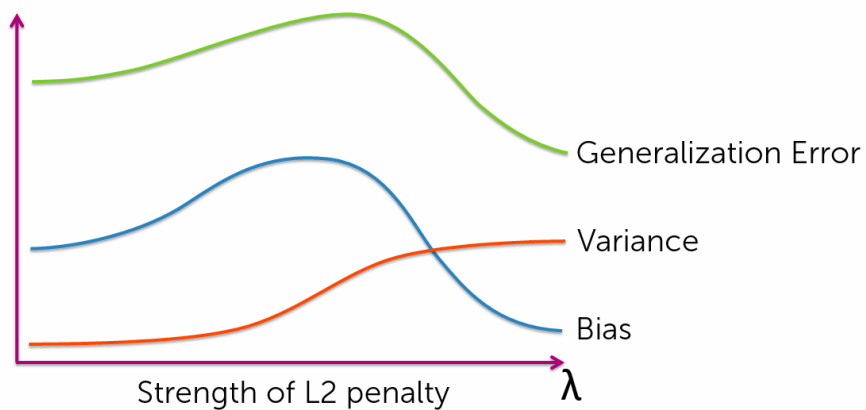
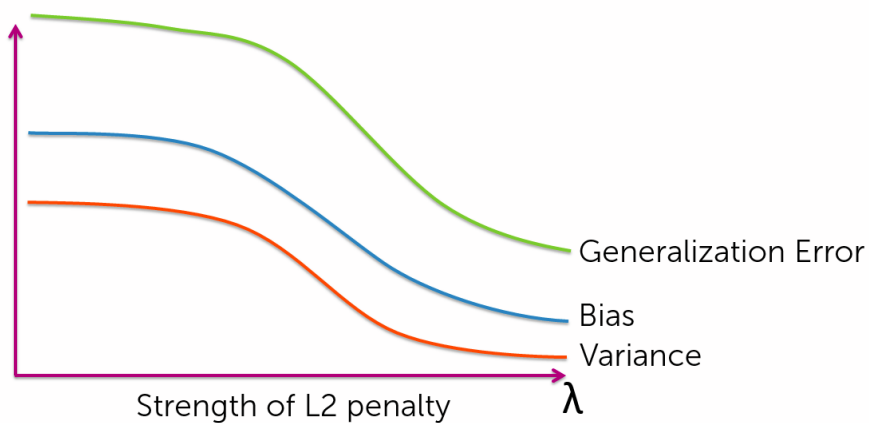
In ridge regression, choosing a large penalty strength λ tends to lead to a model with (choose all that apply):

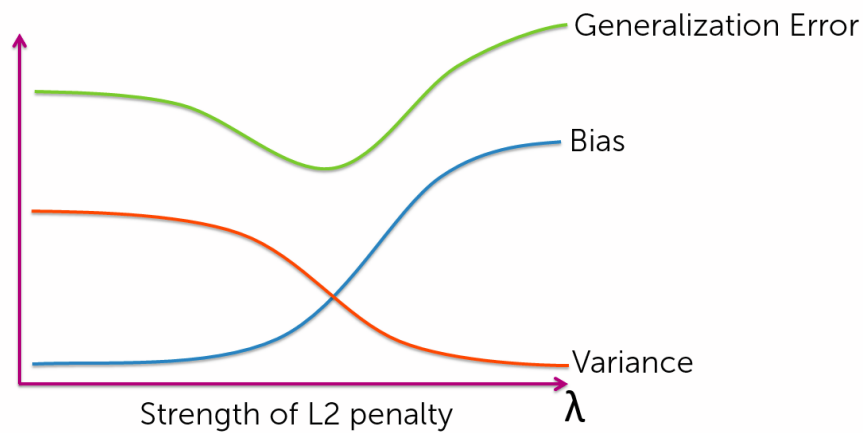
- ☐ High bias
- ☐ Low bias
- ☐ High variance
- ☐ Low variance

1
point

3.

Which of the following plots best characterize the trend of bias, variance, and generalization error (all plotted over λ)?





1
point

4.

In ridge regression using unnormalized features, if you double the value of a given feature (i.e., a specific column of the feature matrix), what happens to the estimated coefficients for every other feature? They:

- ☐ Double
- ☐ Half
- ☐ Stay the same
- ☐ Impossible to tell from the information provided

1
point

5.

If we only have a small number of observations, K-fold cross validation provides a better estimate of the generalization error than the validation set method.

- ☐ True

☒ False

1
point

6.

10-fold cross validation is more computationally intensive than leave-one-out (LOO) cross validation.

☒ True

☐ False

1
point

7.

Assume you have a training dataset consisting of N observations and D features. You use the closed-form solution to fit a multiple linear regression model using ridge regression. To choose the penalty strength λ , you run leave-one-out (LOO) cross validation searching over L values of λ . Let $\text{Cost}(N, D)$ be the computational cost of running ridge regression with N data points and D features. Assume the prediction cost is negligible compared to the computational cost of training the model. Which of the following represents the computational cost of your LOO cross validation procedure?

☒ $L * N * \text{Cost}(N, D)$

☐ $L * N * \text{Cost}(N-1, D)$

☐ $L * D * \text{Cost}(N-1, D)$

☐ $L * D * \text{Cost}(N, D)$

☐ $L * \text{Cost}(N-1, D)$

☐ $L * \text{Cost}(N, D)$

1
point

8.

Assume you have a training dataset consisting of 1 million observations. Suppose running the closed-form solution to fit a multiple linear regression model using ridge regression on this data takes 1 second. Suppose you want to choose the penalty strength λ by searching over 100 possible values. How long will it take to run leave-one-out (LOO) cross-validation for this selection task?

- ☐ About 3 hours
 - ☐ About 3 days
 - ☐ About 3 years
 - ☐ About 3 decades
-

1
point

9.

Assume you have a training dataset consisting of 1 million observations. Suppose running the closed-form solution to fit a multiple linear regression model using ridge regression on this data takes 1 second. Suppose you want to choose the penalty strength λ by searching over 100 possible values. If you only want to spend about 1 hour to select λ , what value of k should you use for k -fold cross-validation?

- ☐ $k=6$
 - ☐ $k=36$
 - ☐ $k=600$
 - ☐ $k=3600$
-

Submit Quiz

