Unit 2: Questions and Exercises

Question 1

For linear models, the two most common methods of fitting the algorithm to the training data are _____ _____ and _____ _____.

Question 2

In supervised learning, the K-class classification problem is viewed as K _____-_____ problems.

Question 3

When the order of the polynomial is increased in linear regression, the error on the _____ data decreases. Higher order polynomials are often too closely fit to a limited set of data points and struggle to generalize to new instances, this is known as _____.

    a. Validation, Overfitting
    b. Validation, Underfitting
    c. Training, Overfitting
    d. Training, Underfitting

Question 4

For a support vector machine, the optimal separating hyperplane is the hypothesis that maximizes the _____

Question 5

When building a model, we should match the _____ complexity with the complexity of the function underlying the data. This is known as the bias-variance tradeoff.

Bonus!

Which Python function from the scikit-learn library can be used to create a validation dataset?

    a. $model\_selection.cross\_val\_score()$
    b. $model\_selection.train\_test\_split()$
    c. $model.predict()$
    d. $dataset.groupby('class').size())$