

1. We have N samples in our training data. What would happen if we wanted to classify test samples using a K -NN classifier with $K=N$? (1 point)

2. What kind of learning is used to generate a decision tree? Supervised, unsupervised, semi-supervised, or reinforcement? Circle one and explain. (2 points)

3. Give an example of noise at the label level. Explain how that example demonstrates label noise. (2 points)

4. According to our discussion in class, how would you map the following categorical values to a feature vector? $\{Dog, Cat, Horse, Bird, Fish\}$ (2 points)

5. Answer the following questions given the training data below. Each sample is listed as $s = (x_1, x_2, y)$, where x_1 and x_2 are the features and y is the label.
 $s_1 = (0, 0, 1)$
 $s_2 = (2, 2, -1)$
 $s_3 = (0, 1, 1)$
 $s_4 = (2, 4, -1)$
 $s_5 = (3, 2, -1)$
 $s_6 = (1, 0, 1)$
 $s_7 = (1, 1, 1)$
 $s_8 = (3, 5, -1)$
 - (a) How would a 1-NN classify the following sample? $s_t = (3, 3)$ (1 point)
 - (b) How would a 3-NN classify the following sample? $s_t = (1, 2)$ (1 point)
 - (c) How would a 5-NN classify the following sample? $s_t = (3, 1)$ (1 point)

1. Why is the K-NN ($K > 1$) classifier better than a 1-NN classifier? (1 point)

2. I generate a Decision Tree using the greedy algorithm from class. My Decision Tree achieves a training accuracy of 90% and a validation accuracy of 60%. Has my model overfit or underfit the training data? Explain. (2 points)

3. Hyperparameters are chosen by analyzing the training data. True or False. Circle one and explain. (2 points)

4. We use the test data to determine the max depth of a decision tree. True or False. Circle one and explain. (2 points)

5. Answer the following questions given the training data below. Each sample is listed as $s = (x_1, x_2, y)$, where x_1 and x_2 are the features and y is the label.
 $s_1 = (0, 0, 1)$
 $s_2 = (2, 2, -1)$
 $s_3 = (0, 1, 1)$
 $s_4 = (2, 4, -1)$
 $s_5 = (3, 2, -1)$
 $s_6 = (1, 0, 1)$
 $s_7 = (1, 1, 1)$
 $s_8 = (3, 5, -1)$
 - (a) How would a 1-NN classify the following sample? $s_t = (3, 3)$ (1 point)
 - (b) How would a 3-NN classify the following sample? $s_t = (1, 2)$ (1 point)
 - (c) How would a 5-NN classify the following sample? $s_t = (3, 1)$ (1 point)