CS491 Quiz 3 Name Solution + Rubric

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

noise: 1-NN subject to noise.

-1 no "noise"

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)

overfit validation should be closer to train accuracy.

-1 wrong explanation

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

use validation parameters are chosen w/ training data, not hyperparameters.

-1 wrong explanation

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

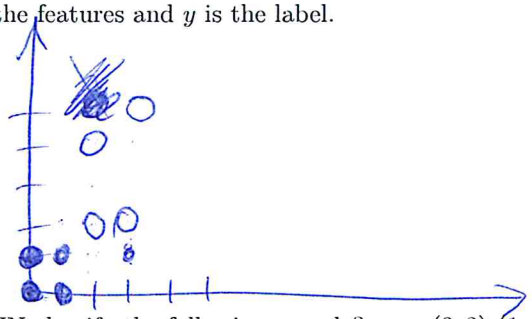
use validation data. dont touch test

-2 true

-1 wrong explanation

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) -1

(b) How would a 3-NN classify the following sample? $s_t = (1, 2)$ (1 point) +1

(c) How would a 5-NN classify the following sample? $s_t = (3, 1)$ (1 point) +1

CS691 Quiz 3 Name Solution + Rubric

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)

Majority class prediction always.

-1 if neither of those underfit

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

Labels are available during training

-1 if no "labels"

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

Good movie review and click thumbs down. Mistake on the label.

-1 bad explanation

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)

Dog = is dog? dog = {1, 0, 0, 0} Cat = {0, 1, 0, 0} etc.

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) -1

(b) How would a 3-NN classify the following sample? $s_t = (1, 2)$ (1 point) +1

(c) How would a 5-NN classify the following sample? $s_t = (3, 1)$ (1 point) +1