

Problem Set 4

1. Consider a data set $(X, Y) = \{(x_i, y_i) : i = 1, \dots, n\}$, where $n > 2$ and all values are positive integers; X is the feature and Y is the class/label. Assume that the data satisfies the following regression model [50]

$$Y = aX + \varepsilon ,$$

where $a \geq 0$ is a deterministic (non-random) parameter and ε is a Bernoulli random variable $B(p)$ (i.e., $\mathbb{P}(\varepsilon = 1) = p$ and $\mathbb{P}(\varepsilon = 0) = 1 - p$).

- (a) Assuming that $a = 0$, apply the method of Maximum-Likelihood-Estimators (MLE) to determine p . [25]
- (b) Assuming that all values x_i are distinct, explain the steps to determine the unknown parameters a and p to minimize the “Residual Sum of (Squared) Errors”; your method must be based on MLE. [25]
2. (a) Consider two points A and B belonging to some class, and two other points C and D belonging to another class. Assume that the points are placed on the circumference of a circle, whereas the distance between any two is the Euclidean distance. Is it possible that the 4-fold cross validation error of k-NN, for $k = 2$, is 1? (Justify!) [15]
- (b) Consider two circles of diameter 1 which are sufficiently apart from each other (the smallest distance between them is at least 1). There are two different points A and B on the circumference of the first circle, and two other different points C and D on the circumference of the other. A and C belong to one class and B and D belong to another class. Assume the following distances: $d(x, y)$ is the Euclidean distance and $d'(y, x) = -d(x, y)$, both defined for $x \in \{A, B\}$ and $y \in \{C, D\}$. Is it possible that the 4-fold cross validation error of k-NN, for $k = 2$, is 1? (Justify!) [35]
-
-