EE 445: Homework Set 1 (DUE 2/9/22)

January 25, 2022

Please let me know if you see a mistake. Thanks!

1 Concept Problems

1. Suppose a discrete random variable X has the following PMFs: $P(1) = 1/2, P(2) = 1/4, P(3) = 1/8, P(4) = 1/8$

   (a) Find and sketch the CDF $F_X(x)$.

   (b) Find $1) P(X \leq 1), 2) P(1 < X \leq 3)$

2. The joint PMF of a bivariate random variable (X, Y) is given by

$$P_{XY}(x, y) = \begin{cases} k(2x_i + y_j), & x_i = 1, 2; y_j = 1, 2 \\ 0, & \text{otherwise.} \end{cases}$$

   where k is a constant.

   (a) Find the value of k

   (b) Find the marginal PMF's of X and Y

   (c) Are X and Y independent

3. A sample of heads and tails is created by tossing a coin a number of times independently. Assume we have a number of coins that generate different samples independently. For a given coin, let the probability of head be μ. The probability of obtaining k heads in N tosses of this coin is given by the binomial distribution

$$p(k|N, \mu) = \binom{N}{k} \mu^k (1 - \mu)^{N-k}.$$

   From the N tosses, the observed probability of head is $\nu = \frac{k}{N}$.

   (a) We have 1 coin. The sample size (N) is 10. If the coin have $\mu = 0.05$, compute the probability that the coin will have $\nu = 0$.

(b) We have 1000 coins. The sample size (N) of each coin is 10. If all the coin have $\mu = 0.05$, , compute the probability that at least one coin will have $\nu = 0$.

(c) For the case $N = 10$ and 1 coin with $\mu = 0.5$. Find and plot the probability

$$P(|\nu - \mu| > \epsilon)$$

for ϵ in the range $[0, 1]$. On the same plot, plot the Hoeffding Inequality $2e^{-2\epsilon^2 N}$.

(d) For the case $N = 6$ and 2 coins with $\mu = 0.5$ for both coins. Find and plot the probability

$$P(\max_{i=1,2} |\nu_i - \mu_i| > \epsilon),$$

for ϵ in the range $[0, 1]$. The index $i = 1, 2$ refers to the two coins. What is the Hoeffding Inequality in this case? On the same plot, plot the Hoeffding Inequality.

(HINT, $P(max(A, B) > \epsilon) = P(A > \epsilon)$ OR $P(B > \epsilon) = P(A > \epsilon) + P(B > \epsilon) - P(A > \epsilon, B > \epsilon)$)

(e) Explain the relationship between the coins, probability of heads, etc. and how they are related to the concepts of target functions, errors, hypothesis, etc.

4. This problem investigates how changing the error measurement can change the result of the learning process. You have N data points $y_1 \leq y_2 \leq \ldots \leq y_N$ and wish to estimate a representative value.

(a) If your algorithm is to find the hypothesis h that minimizes the in-sample sum of squared deviations

$$E_{in}(h) = \sum_{n=1}^{N} (h - y_n)^2,$$

then show that your estimate will be the in-sample mean,

$$h = \frac{1}{N} \sum_{n=1}^{N} y_n.$$

(HINT: For a quadratic formula, $ax^2 + bx + c$, if $a > 0$, then the minimum point occurs at $x = \frac{-b}{2a}$)

(b) If your algorithm is to find the hypothesis h that minimizes the in-sample sum of absolute deviations

$$E_{in}(h) = \sum_{n=1}^{N} |h - y_n|,$$

then show that your estimate will be the in-sample median. The median point is any value for which half the data points are at most h_{median} and half the data points are at least h_{median}.

(c) Suppose one of the data points, say y_N, is perturbed to $y_N + \epsilon$, where ϵ is a large number. This point then becomes an outlier. What happens to your hypotheses?

2 Coding Problems

1. (a) You are given a linearly separable dataset of size 20 in 20sample_d2.csv, where the input space is $\mathcal{X} = \mathcal{R}^2$ ($d = 2$). Plot the examples $\{(\mathbf{x}_n, y_n)\}$ as well as the target function. Be sure to mark the examples from different classes differently and add labels to the axes of the plot. The data was artificially generated. The true target function is $w = [0.1860, 0.2654]^T, b = 0$.

   (b) Run the perceptron learning algorithm on 20sample_d2.csv. Report the number of updates that the algorithm takes before converging. Comment on whether the target function f is close to g.

   (c) Repeat everything in (b) with another randomly generated generated data set of size $N = 20$, 2_20sample_d2.csv. Compare your results with (b). The true target function is $w = [0.1245, 0.9486]^T, b = 0$

   (d) Repeat everything in (b) with another randomly generated data set of size $N = 1000$, 1000sampe_d2.csv. Compare your results with (b). The true target function is $w = [-0.9469, -0.1828]^T, b = 0$

   (e) We now use a dataset where the input space is $\mathcal{X} = \mathcal{R}^{10}$ ($d = 10$), 1000sample_d10.csv. Now the points can no longer be visualized. Run the perceptron learning algorithm on where $N = 1000$. How many updates does the algorithm take to converge?

   (f) Repeat the algorithm on the same data set as (e) for 100 experiments. For each experiments, randomly select your misclassified points. Plot a histogram for the number of updates that the algorithm takes to converge

   (g) Summarize your conclusions with respect to accuracy and running time as a function of N and d.