

Homework 2

Problem 1

b is the correct answer.

See attached code. After running the simulation, the average value of v_{min} was 0.03, closest to 0.01 or answer **b**.

Problem 2

d is the correct answer.

See attached code. My values for v_1 and v_{rand} were 0.49 and 0.48, respectively. These values are close to 0.5 which is equal to μ . Thus, the corresponding coins of c_1 and c_{rand} satisfy the single-bin Hoeffding Inequality.

Problem 3

e is the correct answer.

There are two ways of making an error.

1: Hypothesis h makes an error approximating the function when $y = f(x)$ which happens at a probability of $(\lambda) * \mu$.

2: Hypothesis h does not make an error in approximation and $y \neq f(x)$ which happens at a probability of $(1-\lambda) * (1-\mu)$.

Thus, the overall probability is the sum which is $(\lambda) * \mu + (1-\lambda) * (1-\mu)$.

Problem 4

b is the correct answer.

When $\lambda = 0.5$, $y = f(x)$ happens at the same probability as $y \neq f(x)$. Thus, the performance of h does not depend on $P(y | x)$ and so independence ensues.

Problem 5

c is the correct answer.

See attached code. Using the linear regression technique to calculate the weight vector ($w = X^+ \cdot y$) and calculating E_{in} resulted in a value of 0.02, closest to answer choice **c**.

Problem 6

c is the correct answer.

See attached code. Using the weight vector from problem 5 generated from 100 initial training points and finding the probability of 1000 fresh points for out-of-sample error led to an output of 0.025, closest to choice **c**.

Problem 7

a is the correct answer.

See attached code. For each of the 1000 runs, 10 initial training points were taken for the weight vector to be based off by linear regression. Once this was done, the PLA was ran with this weight vector as the starting weight vector and average number of iterations over 1000 runs was calculated to be 3.411, closest to choice a.

Problem 8

d is the correct answer.

See attached code. A set of 1000 training points were generated with approximately 1/10 of them with their sign flipped. The weight vector was found from these training points using linear regression and the in-sample error was then calculated to be 0.5 as averaged out over 1000 runs, closest to answer choice d.

Problem 9

a is the correct answer.

See attached code. A set of 1000 training points were used per trial (total of 1000 trials). The weights vector was found for each trial and the individual elements of the vector were summed over 1000 trials and averaged out to generate the "coefficient." Out of all the answers, the coefficients generated most matched that of answer choice a.

Problem 10

b is the correct answer.

Using the weight vectors found from problem 9, out of sample error was calculated with 1000 new points with noise for each trial. Over 1000 trials, the average out of sample error was found to be 0.13, closest to answer choice b.