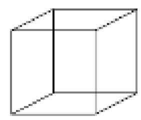
Ariel University
Machine Learning
Homework 1

Problem 1. In this problem we consider the infinite set of 3-dimensional (axis-parallel) rectangles which label points inside as + and points outside as -.



a) Give a PAC bound for this set without using VC-theory.
b) What is the VC-dimension of this set? Prove your answer.
c) Give a PAC bound for this set using VC-theory.
d) Compare the two bounds.

Problem 2. In the game of Risk, player **A** attacks player **B** over many rounds. At each round, **A** rolls two dice, and **B** rolls one die. Each die has six sides (and values in {1, 2, 3, 4, 5, 6}) and is fair. If one or both of **A**'s dice rolls is <u>greater</u> than **B**'s, then **A** wins the round. Otherwise **B** wins the round.

a) What's the probability that **A** wins a given round? What's the probability that **B** wins a given round? If there are 100 rounds, what is the expected number of rounds that **A** and **B** each win?
b) If there are 100 rounds, give an upper bound on the probability that **A** will win 65 or more rounds. Derive bounds using Markov and Hoeffding.
c) At least how many rounds do **A**,**B** need to play so that with at least 90% probability, **B** will win 35 or more rounds? Derive bounds using Hoeffding.

Problem 3. Let our set of rules H include all programs of size exactly 100 bits. Suppose we sampled n points from a distribution D and found that one of the programs is consistent with 80% of the sample points.

a) What is the smallest δ for which we can say that: With probability at least $1-\delta$, the program misclassifies at most a $.2+\epsilon$ fraction of all points of D? (δ is a function of n,ϵ.)
b) Now suppose n=1000. Give a minimum value ϵ such that with probability at least 90%, the program misclassifies at most a $.2+\epsilon$ fraction of all points of D.
c) In problem (b) above, what can you say about the true error of the program on points of a different distribution D'?