UC Berkeley
Department of Electrical Engineering and Computer Science
Department of Statistics

EECS 281A / STAT 241A Statistical Learning Theory

Problem Set 1

Fall 2016

Issued: Tues, August 30, 2016 **Due:** Thursday, September 8, 2016

Comments: The first five problems are a combination of standard undergraduate probability/statistics and linear algebra. If you don't understand the statements or find these problems very challenging, it is a diagnostic sign that you lack the appropriate background for taking this class.

Problem 1.1

Hana and Owen love to challenge each other to coin flipping contests. Suppose that Owen flips $n+1$ coins, and Hana flips with n coins, and that all coins are fair. Letting E be the event that Owen flips more heads than Hana, show that $\mathbb{P}[E] = \frac{1}{2}$.

Problem 1.2

A group of N archers shoot at a target. The distance of each shot from the center of the target is uniformly distributed between 0 to 1, independently of the other shots. The winner's arrow is closest to the origin, whereas the loser's arrow is farthest away.

(a) Find the expected distance from the winner's arrow to the center.

(b) Find the expected distance from the loser's arrow to the center.

Problem 1.3

Every day that he leaves work, Albert the Absent-minded Professor toggles his light switch according to the following protocol: (i) if the light is on, he switches it off with probability 0.60; and (ii) if the light is off, he switches it on with probability 0.20. At no other time (other than the end of each day) is the light switch touched.

(a) Suppose that on Monday night, Albert's office is equally likely to be light or dark. What is the probability that his office will be lit all five nights of the week (Monday through Friday)?

(b) Suppose that you observe that his office is lit on both Monday and Friday nights. Compute the expected number of nights, from that Monday through Friday, that his office is lit.

Now suppose that Albert has been working for five years (i.e., assume that the Markov chain is in steady state).

(c) Is his light more likely to be on or off at the end of a given workday?

Problem 1.4
A pair of symmetric matrices P and Q is simultaneously diagonalizable if there is an orthonormal matrix U such that U^TPU and U^TQU are both diagonal. Show that this happens if and only if $PQ = QP$.

Problem 1.5
The test scores of 900 students had the following sample statistics: Mean 83 and Variance 36.

(a) Suppose that we select one student uniformly at random. Use Chebyshev's inequality to bound the probability that this randomly selected student received a test score between 71 and 95 inclusive.

(b) Do you think that it likely that at least 600 students scored between 71 and 95 inclusive? Why or why not? Explain your reasoning.

Problem 1.6
Linear regression and gradient descent: Given a data matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ and a response vector $y \in \mathbb{R}^n$, consider the least-squares method for constructing a linear predictor $x \mapsto \langle \widehat{\theta}, x\rangle$:

$$\widehat{\theta} = \arg\min_{\theta\in\mathbb{R}^d} \left\{ \frac{1}{2n} \|y - \mathbf{X}\theta\|_2^2 \right\}.$$

(a) Under what conditions is the solution $\widehat{\theta}$ unique?

(b) Assuming that $\mathbf{X}^T\mathbf{X}$ is invertible, show that $\widehat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^Ty$.

(c) Write some code (in MATLAB, R, Python etc.) to implement the standard gradient descent updates with an arbitrary stepsize $\alpha > 0$. Using the $(\mathbf{X}, y)$ data in `RegressionData.zip`, experiment with your code to see for what values of α it converges or fails to converge.

(d) Prove that gradient descent converges for all stepsizes α in an interval of the form $(0, C)$ for some constant $C > 0$. Try to find the largest possible C. (*Hint:* Your answer should involve eigenvalues of the matrix $\mathbf{X}^T\mathbf{X}$.)