CS446: Machine Learning **Fall 2016**

Problem Set 0

Handed Out: August 23^{rd}, 2016 *Due: NONE*

- Feel free to talk to other members of the class in doing the homework. I am more concerned that you learn how to solve the problem than that you demonstrate that you solved it entirely on your own. You should, however, write down your solution yourself. Please try to keep the solution brief and clear.
- Feel free to send me email or come to ask questions.
- You should have seen all the material; the goal of this homework is to allow you to go back to some of this and refresh your memory. Consequently, you do not need to turn in this homework; we do hope that you do it, write it down (only then you will make sure you understand) and compare your solutions to the solutions we will supply next week.

1. [Probability] Assume that the probability of obtaining heads when tossing a coin is λ.
   a. What is the probability of obtaining the first head at the $(k+1)$-th toss?
   b. What is the expected number of tosses needed to get the first head?
2. [Probability] Assume X is a random variable.
   a. We define the variance of X as: $Var(X) = E[(X-E[X])^2]$. Prove that $Var(X) = E[X^2] - E[X]^2$.
   b. If $E[X] = 0$ and $E[X^2] = 1$, what is the variance of X? If $Y = a + bX$, what is the variance of Y?
3. [Probability] John is a great fortune teller. Assume that we know three facts: 1) If John tells you that a lottery ticket will win, it will win with probability 0.99. 2) If John tells you that a lottery ticket will not win, it will not win with probability 0.99999. 3) With probability 10^{-5}, John predicts that a ticket is a winning ticket. This also means that with probability $1 - 10^{-5}$, John predicts that a ticket will not win.
   a. Given a ticket, what is the probability that it wins?
   b. What is the probability that John correctly predicts a winning ticket?
4. [Calculus] Let $f(x, y) = 3x^2 + y^2 - xy - 11x$
   a. Find $\frac{\partial f}{\partial x}$, the partial derivative of f with respect to x. Find $\frac{\partial f}{\partial y}$.
   b. Find $(x, y) \in \mathbb{R}^2$ that minimizes f.
5. [Linear Alegbra] Assume that $w \in \mathbb{R}^n$ and b is a scalar. A hyper-plane in R^n is the set, $\{x : x \in \mathbb{R}^n, w^T x + b = 0\}$.
   a. For $n = 2$ and 3, find two example hyper-planes (say, for $n = 2$, $w^T = \begin{bmatrix}1 & 1\end{bmatrix}$ and $b = 2$ and for $n = 3$, $w^T = \begin{bmatrix}1 & 1 & 1\end{bmatrix}$ and $b = 3$) and draw them on paper.

b. The distance between a point $x_0 \in \mathbb{R}^n$ and the hyperplane $w^T x + b = 0$ can be described as the solution of the following optimization problem:

$$\begin{aligned} &\min_x \|x_0 - x\|^2 \\ &\text{s.t. } w^T x + b = 0 \end{aligned}$$

However, it turns out that the distance between x_0 and $w^T x + b = 0$ has an analytic solution. Derive the solution. (*Hint: you may be familiar with another way of deriving this distance; try your way too.*)

c. Assume that we have two hyper-planes, $w^T x + b_1 = 0$ and $w^T x + b_2 = 0$. What is the distance between these two hyperplanes?

6. [Linear Algebra] One way to define a *convex* function is as follows. A function $f(x)$ is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for all x, y and $0 < \lambda < 1$.

a. Prove that $f(x) = x^2$ is a convex function. (Prove by applying the definition.)

b. A n-by-n matrix A is a *positive semi-definite* matrix if $x^T A x \geq 0$, for any $x \in \mathbb{R}^n$ s.t $x \neq 0$.

Prove that the function $f(x) = x^T A x$ is convex if A is a positive semi-definite matrix. Note that x is a vector here. (*Hint: the solution is somewhat similar to the solution of part (a.)*)

7. [CNF and DNF] Consider the following Boolean function written in a conjunctive normal form

$$(x_1 \vee x_2) \wedge (x_3 \vee x_4) \wedge \ldots (x_{15} \vee x_{16})$$

If no new variable is introduced, how many clauses do you need to write down the same function in disjunctive normal form ?