**CIS 419/519: Applied Machine Learning** **Fall 2020**

# Homework 0

*Handed Out: September 2* *Due: September 14*

- Name: Tangji Li
- Pennid: 76947215
- The goal of HW0 is to give you an idea of the level of mathematical knowledge and maturity expected in this class. You should have seen all the material used here; the goal of this homework is to encourage you to go back to some of the material and refresh your memory.
- Most of these problems require you to be familiar with the notation and the definitions of the concepts. Beyond that, you should be able to reason from the definitions and do some algebraic manipulation. None of these problems require deep understanding of the material.
- Feel free to go back to some of your text books and the web when solving it. You should, however, be able to write down your own solution to these problems.
- **If you are not familiar with more than 30% of the material in this problem set, or find it difficult to solve at least 2/3 of it in a few hours, you are missing the prerequisites, and you will find the class too difficult. In this case, I recommend that you drop the class, take the required classes first, and then come back to take this class.**
- While we encourage discussion within and outside the class, in the specific case of HW0, you are on your own. It's a test of your level of readiness to the clas, so please work on it independently.
- Please use Piazza if you have questions about the homework. In general, we encourage you to come to the Professor and the TAs office ours and recitations, although for this specific assignment you are on your own.
- Please try to keep the solution brief and clear.
- Please, no handwritten solutions. Consult the class' website if you need guidance on using Latex. You will submit your solution manuscript as a single pdf file.
- The homework is due at 11:59 PM on the due date. We will be using Canvas for collecting the homework assignments. Please submit your solution manuscript as a pdf file via Canvas. Please do NOT hand in a hard copy of your write-up. Post on Piazza and Contact the TAs if you are having technical difficulties in submitting the assignment.
- **If you are on the waiting list**, we recommend that you still work on the assignment so that you can determine if you want to stay on the waiting list. If you cannot register for the class by the time the assignment is due, please email your solutions to Ben Zhou (`xyzhou@seas.upenn.edu`) with "CIS 419/519 Homework 0" in the subject.

1. [Probability] Assume that the probability of obtaining heads when tossing a coin is $\lambda$.

   a. What is the probability of obtaining the first head at the $(k+2)$-th toss?

   b. Assume the $(n)$-th and $(n+1)$-th tosses are heads and all previous tosses are tails, what is $n$'s expectation?

   - Solution: a. $(1-\lambda)^{k+1}\lambda$
     b. assuming we get first head of coin needs $T_{n-1}$, then for the second coin of head, we need $T_n = T_{n-1} + 1 + \lambda * 0 + (1 - \lambda)T_n$, since for the $T_1$, we got $T_1 = \frac{1}{\lambda}$, we can easily get $T_n = \frac{1}{\lambda - 1}(1 - \frac{1}{\lambda^n})$

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 $Y$?

   (a) Solution: a. from the definition of the variance:

$$E[(X - E[X])^2 = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[XE[X]] + E[X]^2$$
$$= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2 \qquad (1)$$

b. From Eq.1, when $E[X] = 0$ and $E[X^2] = 1$, we can easily get $Var(X) = E[X^2] - E[X]^2 = 1$
if $Y = a - bX$, then

$$E[Y^2] = E[(a - bX)^2] = E[a^2 - 2abX + b^2Y^2]$$
$$= a^2 - 2abE[X] + b^2E[X^2] = a^2 + b^2$$
$$E[Y] = E[a - bX] = a - bE[X] = a$$
$$Var(Y) = E[Y^2] - E[Y]^2 = b^2$$

3. [Probability] John is a great fortune teller. Assume that we know three facts:

- If John tells you that a lottery ticket will win, it will win with probability 0.98.
- If John tells you that a lottery ticket will not win, it will not win with probability 0.99.
- When John sees a ticket he predicts with probability $10^{-5}$ 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?

(a) Solution: a. let M be the event "John predicts a ticket is winning ticket", let N be the event "the ticket wins"
then

$$P(N) = P(M,N) + P(\neg M,N) = P(N|M)P(M) + P(N|\neg M)P(\neg M)$$
$$= 0.98 \times 10^{-5} + (1 - 0.99) \times (1 - 10^{-5})$$
$$= 0.0100097$$

b.The probability John predicts correct win is:

$$P(M|N) = \frac{P(M,N)}{P(N)} = \frac{P(N|M)P(M)}{P(N)}$$

$$= \frac{0.98 \times 10^{-5}}{0.98 \times 10^{-5} + (1-0.99) \times (1-10^{-5})}$$
$$\approx 0.000979$$

4. [Calculus] Let $f(x,y) = x^2 + 2y^2 - 2xy - x$

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 or maximizes $f$ and show whether it is a global minimum or maximum.

- Solution: a.$\frac{\partial f}{\partial x} = 2x - 2y - 1$
$\frac{\partial f}{\partial y} = 4y - 2x$

  b.let both partial equation equals 0, we can get the stationary point$(1, 1/2)$ and for the second derivative, we can get

$$A = f_{xx} = 2$$

$$B = f_{xy} = -2$$

$$C = f_{yy} = 4$$

then $H_f = 8 - 4 = 4 > 0$, it is positive definite, and A is positive, so it is the minimum point. Consider for the boundary and infinite point, it is positive, so the point(1,1/2) is the global minimum point.

5. [Linear Algebra] 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 a 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:

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

   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.*)

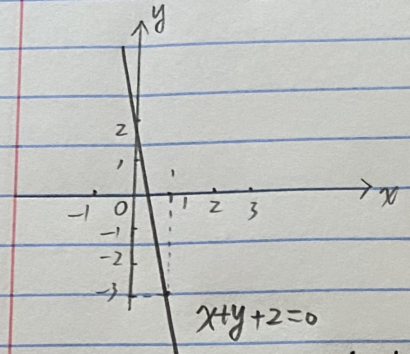
   c. Assume that we have two hyper-planes, $w^T x + b_1 = 0$ and $w^T x + b_2 = 0$. What is the distance between these two hyperplanes?

CIS 519 ML Name: Tangji Li Pennid: 76947215

5.

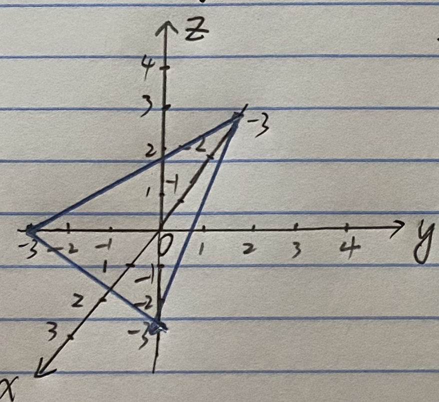
a. for $n=2$, it is a 2-dplane.

take $x+y+2=0$ as an example.




for $n=3$, it is 3d plane.

take $x+y+z+3=0$ as an example.




the solution is a 2-d plane. (drawed by blue pen).

b. In an n-dimensional space, let $\vec{x}_1$ be the point in hyperplane $\vec{w}^T\vec{x}+b=0$. and $\min_{x} \|\vec{x}_0-\vec{x}\|^2$

so $\vec{x}_0-\vec{x}$ is orthogonal to the plane.

So $\vec{x}_0-\vec{x} = \|\vec{x}_0-\vec{x}\| \frac{\vec{w}}{\|\vec{w}\|}$

so $\vec{w}^T\vec{x}_0 - \vec{w}^T\vec{x} = \|\vec{x}_0-\vec{x}\| \frac{\|\vec{w}\|^2}{\|\vec{w}\|}$

since $x_1$ is on the $\vec{w}^Tx+b=0$, so $-\vec{w}^T\vec{x}=b$.

so $\|\vec{x}_0-\vec{x}\| = \frac{|\vec{w}^T\vec{x}_0+b|}{\|\vec{w}\|}$

c. fetch a point on plane 1, we have $\vec{w}^T\vec{x}_0+b_1=0$

use the b. eq. the distance $= \frac{|b_2-b_1|}{\|\vec{w}\|}$

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. Determine which of the following functions are convex functions?

   i. $(x+1)^2$

   ii. $g(x) = f_1 * f_2$ where $f_1$,$f_2$ are given to be convex functions.

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

   Use the definitions to show that the function $f(x) = x^T Ax$ 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.)*)

(a) Solution: a. for i. use the definition of convex function, we can get

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

$$= (\lambda x + (1-\lambda)y + 1)^2 - \lambda(x+1)^2 - (1-\lambda)(y+1)^2$$

$$= (\lambda - 1)\lambda(x-y)^2 <= 0$$

because $0 < \lambda < 1$.

for ii. also by definition, since $f_1$ and $f_2$ are convex function, then the times of both will $> 0$, so $g(x)$ will not be a convex function.

b. use the definition, we can get

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

$$= (\lambda x + (1-\lambda)y)^T A(\lambda x + (1-\lambda)y) - \lambda x^T Ax - (1-\lambda)y^T Ay$$

$$= (\lambda - 1)\lambda(x^T Ax + y^T Ay - x^T Ay - y^T Ax)$$

$$= (\lambda - 1)\lambda(x-y)^T A(x-y) <= 0$$

because A is positive semi-definite.

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

$$(x_1 \vee x_2 \vee \neg x_3) \wedge x_4$$

Convert it to the disjunctive normal form.

1. Solution: use the distributive laws, we can get

$$(x_1 \wedge x_4) \vee (x_2 \wedge x_4) \vee (\neg x_3 \wedge x_4)$$