COMS 4771 HW0

Due: Sun Sept 09, 2018 at 11:59pm

This is a calibration assignment (HW0). The goal of this assignment is for you to recall basic concepts, and get familiarized with the homework submission system (Gradescope). Everyone on the waitlist intending to enroll must submit this assignment by the due date. Anyone who does not submit HW0 by the due date will get a score of zero. The score received on this assignment will not count towards your final grade in this course, but will be used to make a decision to who will be approved to enrolled . You must show your work to receive full credit.

This homework assignment is to be done individually. All homeworks (including this one) should be typesetted properly in pdf format. Handwritten solutions will not be accepted. You must include your name and UNI in your homework submission.

1.1 **[Probability and Statistics]** Let X and Y be discrete random variables, and consider the joint distribution (X, Y) given by

	Y=1	Y=2	Y=3
X=1	0.2	0.2	0.3
X=2	0.1	0.1	0.1

(i) What is the marginal distribution of X?

(ii) What is $\Pr[Y = 1|X = 2]$?

(iii) Let $f : x \mapsto x^2$. What is $\mathbb{E}[f(X)|Y = 3]$?

1.2 Fix some $\theta > 0$, and consider the function $g_\theta : [0, \infty) \to \mathbb{R}$ defined as $x \mapsto \frac{1}{\theta} e^{-\frac{x}{\theta}}$.

(i) Verify that g_θ is a probability distribution.

(ii) What is $\mathbb{E}[g_\theta]$?

(iii) What is Variance(g_θ)?

1.3 Let X and Y be jointly distributed Gaussian random variables, where

$$\mathbb{E}(X) = 1, \mathbb{E}(Y) = -1,$$
$$\text{var}(X) = 1, \text{var}(Y) = 9,$$
$$\text{cov}(X, Y) = -2$$

What is the distribution of $X + Y$?

1.4 If you independently toss a fair coin n times, what is the expected absolute difference between the number of heads H and the number of tails T? In other words, what is $\mathbb{E}\big[|H - T|\big]$?

2.1 **[Linear Algebra]** Consider the subspace S spanned by vectors $\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 10 \\ 6 \\ 6 \end{pmatrix}$ in $\mathbb{R}^4$.

(i) What is the dimension of the subspace S.

(ii) Compute the orthogonal linear projection of the point $\begin{pmatrix} 6 \\ 5 \\ 9 \\ 2 \end{pmatrix}$, onto the subspace S.

2.2 Prove that for any $m \times n$ real matrix $\mathbf{A}$ and any $\rho > 0$, the matrix $\mathbf{A}^\mathsf{T}\mathbf{A} + \rho\mathbf{I}$ is invertible (where $\mathbf{I}$ is the $n \times n$ identity matrix). (Hint: show that all eigenvalues of $\mathbf{A}^\mathsf{T}\mathbf{A} + \rho\mathbf{I}$ are real and analyze the smallest eigenvalue.)

3.1 **[Calculus and optimization]** Let $\mathbf{A} \in \mathbb{R}^{d \times D}$ be a real matrix, and $\mathbf{b} \in \mathbb{R}^d$ be a real vector. Define the function

$$\begin{aligned} f : \mathbb{R}^D &\to \mathbb{R} \\ \mathbf{x} &\mapsto \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \|\mathbf{x}\|^2. \end{aligned}$$

(i) What is $\nabla f(\mathbf{x})$?

(ii) What value of $\mathbf{x}$ minimizes f, that is, find $\arg\min_{\mathbf{x}} f(\mathbf{x})$? (Hint: compute the stationary points of f.)

3.2 Let $g : \mathbb{R}^2 \to \mathbb{R}$ defined as $g(x) := x^\mathsf{T} A x - b^\mathsf{T} x + c$, where $A := \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$, $b := \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and $c := 3$. What is $\nabla g((1, 1))$?

4.1 **[Programming practice]** Download the Matlab data file `hw0data.mat` (instructions on Piazza on where to download the file). Write a script that does the following.

Special note for those who are not using Matlab: Python users can use `scipy` to read in the mat file, R users can use `R.matlab` package to read in the mat file, Julia users can use `JuliaIO/MAT.jl`. Octave users should be able to load the file directly.

(i) Load the data in `hw0data.mat`. It contains one matrix variable is called $\mathbf{M}$.

(ii) Print the dimensions of $\mathbf{M}$.

(iii) Print the 4th row and 5th column entry of $\mathbf{M}$.

(iv) Print the mean value of the 5th column of $\mathbf{M}$.

(v) Compute the histogram of the 4th row of $\mathbf{M}$ and show the figure.

(vi) Compute and print the top three eigenvalues of the matrix $\mathbf{M}^\mathsf{T}\mathbf{M}$.

4.2 We will try to understand the geometry of eigenvectors and eigenvalues of a matrix via experimentation. Let $\mathbf{L} = \begin{bmatrix} 5/4 & -3/2 \\ -3/2 & 5 \end{bmatrix}$ be a 2×2 matrix. To understand eigenvectors and eigenvalues, we will study the *action* of L on random vectors and relate it to eigenvectors and eigenvalues. Write a script that does the following.

(i) Create the 2×2 matrix $\mathbf{L}$ (as defined above).

(ii) Create 500 random, unit length, two-dimensional vectors. (Hint: to generate a random d-dimensional unit length vector, draw d independent samples from the Gaussian distribution $N(0, 1)$ and assign each sample as one component of the vector. Now, normalize the vector to have length one.) Let R be the set of these 500 random 2-dimensional unit vectors.

(iii) For each vector $\mathbf{r} \in R$, compute how the matrix $\mathbf{L}$ "distorts" $\mathbf{r}$, that is, compute $\tilde{\mathbf{r}} := \mathbf{L}\mathbf{r}$.

(iv) Compute the eigenvalues of $\mathbf{L}$. Let $\lambda_{\max}$ and $\lambda_{\min}$ denote the maximum and the minimum eigenvalue respectively.

(v) For each distorted vector $\tilde{\mathbf{r}}$, compute the length $\|\tilde{\mathbf{r}}\|$.

(vi) Create a histogram of values of $\|\tilde{\mathbf{r}}\|$ (use 50 bins) and compare it to $\lambda_{\max}$ and $\lambda_{\min}$.

(vii) What relationship can you infer between $\|\tilde{\mathbf{r}}\|$, $\lambda_{\max}$ and $\lambda_{\min}$?

(viii) Now, compute the eigenvectors of $\mathbf{L}$. Let $\mathbf{v}_{\max}$ denote the eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}$.

(ix) Make a two-dimensional plot of all the distorted vectors $\tilde{\mathbf{r}}$ (in blue color) and the eigenvector $\mathbf{L}\mathbf{v}_{\max}$ (in red color). (make sure that the x- and the y-axis are displayed at the same scale).

(x) What can you infer about the $\mathbf{v}_{\max}$ from studying this plot?