COMS 4771 HW0

Due: Sun Jan 27, 2019 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 enrolled or 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. You should cite all resources (including online material, books, articles, help taken from specific individuals, etc.) you used to complete your work.

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.

0.1 **[Notation]**

- $\Pr[\cdot]$ denotes the probability (of an event).
- $\mathbb{E}[\cdot]$ denotes the expected value (of a random variable).
- $\text{var}[\cdot]$ denotes the variance (of a random variable).
- $\text{cov}[\cdot,\cdot]$ denotes the covariance (between a pair of random variables).
- $\mathbf{1}[\cdot]$ denotes the indicator function. That is, $\mathbf{1}[A] := \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases}$.
- $\perp$ denotes independence. That is, $A \perp B$ means A and B are independent.
- $^\mathsf{T}$ denotes the transpose operator.
- $\|\cdot\|$ denotes the Euclidean norm.

1.1 **[Probability and Statistics]**

Let X and Y be jointly distributed normal random variables, where

$$\begin{aligned} \mathbb{E}[X] &= 1, \qquad \mathbb{E}[Y] = -1, \\ \text{var}[X] &= 1, \qquad \text{var}[Y] = 9, \\ \text{cov}[X,Y] &= -2. \end{aligned}$$

In other words, the joint distribution of the pair $(X,Y) \sim N(\mu, \Sigma)$, where

$$\mu := \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \text{ and } \Sigma := \begin{bmatrix} 1 & -2 \\ -2 & 9 \end{bmatrix}.$$

What is the distribution of the random variable $Z := X - 2Y$?

1.2 Show that:

(i) For any (measurable) event A, we have $\Pr[A] = \mathbb{E}[\mathbf{1}[A]]$.

(ii) For any non-negative random variable X, and any $c > 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}.$$

(Hint: compare the output of the function $\mathbf{1}[X > c]$ with the outcome of X.)

1.3 Let $X \in \{-1, +1\}$ denote the outcome of an toss of an unbiased coin. (That is, $\Pr[X = +1] = \Pr[X = -1] = 1/2$.) Say the coin in tossed 1000 times independently, and the correspoinding outcomes are denoted by $X_1, \ldots, X_{1000}$.

Give a good estimate of the chance that the average of the 1000 tosses exceeds the value 10? That is, give the best possible value of α, such that

$$\Pr[(X_1 + \ldots + X_{1000})/1000 > 10] < \alpha.$$

(Hint: use the result from q1.2.)

1.4 Suppose x is a random vector drawn from a d-dimensional multivariate Gassian distribution with mean 0 and covariance Σ. Define $y := Qx + v$, for a known (invertible) $d \times d$ matrix Q, and a $d \times 1$ vector v. What is the distribution of y?

1.5 For any three random variables A, B and C prove or disprove the following statements:

(i) $(A \perp B) | C \implies A \perp B$.

(ii) $A \perp B \implies (A \perp B) | C$.

2.1 **[Linear Algebra]** Let v be a vector in $\mathbb{R}^d$. Consider the set $S_v := \{x \in \mathbb{R}^d \mid x \cdot v = 0\}$. What is the dimension of S_v? (Justify your answer).

2.2 Let S_v be as defined in q2.1, and w be another vector in $\mathbb{R}^d$. What is the Euclidean distance between w and the closest point to w in S_v? That is, find

$$\min_{x \in S_v} \|w - x\|.$$

(for those who are curious, this minimum exists and you don't need to prove its existance.)

2.3 Show that for any symmetric positive semi-definite $d \times d$ real matrix A, there exists real vectors $v_1, \ldots, v_d$ such that

$$A = \sum_{i=1}^{d} v_i v_i^\mathsf{T}.$$

(Hint: consider the eigendecomposition of A)

3.1 **[Calculus and optimization]** For a given vector $b \in \mathbb{R}^d$, Consider the function $F_b : (\mathbb{R}^{d\times d} \times \mathbb{R}^d) \to \mathbb{R}$ defined as:

$$F_b : (A, v) \mapsto v^\mathsf{T} A v + b^\mathsf{T} b.$$

(i) What is $\partial F_b / \partial A$?

(ii) What is $\partial F_b / \partial v$?

(iii) For a fixed invertible and symmetric matrix A, what value of v minimizes the function F_b?

(iv) For a fixed invertible and symmetric matrix A, what value of v such that $\|v\| = 1$ minimizes the function F_b?

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 **M**.

(ii) Print the dimensions of **M**.

(iii) Print the 4th row and 5th column entry of **M**.

(iv) Print the mean value of the 5th column of **M**.

(v) Compute the histogram of the 4th row of **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 **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 **L** "distorts" **r**, that is, compute $\tilde{\mathbf{r}} := \mathbf{L}\mathbf{r}$.

(iv) Compute the eigenvalues of **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 black 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?