

COMS 4721/4771 HW0 (Spring 2024)

No Due Date

This is a calibration assignment (HW0). The goal of this assignment is for you to recall basic concepts, and self-assess your preparedness with the prerequisite knowledge that is expected from you for 4721/4771. HW0 will not be graded, nor will it count towards your final score for this class. Instead, you can self-check the correctness of your work with the solutions provided.

You will be expected to know and be comfortable working with this and similar type of background math and programming knowledge for COMS 4721/4771.

1 Probability and Statistics

- (a) 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 $\mathbb{P}[Y = 1 | X = 2]$?
- (iii) Let $f : x \mapsto x^2$. What is $\mathbb{E}[f(X) | Y = 3]$?
- (b) Fix some $\theta > 0$, and consider the function $g_\theta : [0, \infty) \rightarrow \mathbb{R}$ defined as $x \mapsto \frac{1}{\theta} e^{-\frac{x}{\theta}}$.
- (i) Verify that g_θ is a probability distribution.

Consider the random variable $X \sim g_\theta$.

- (ii) What is $\mathbb{E}[X]$?
- (iii) What is $\text{Var}[X]$?
- (c) Let X and Y be jointly distributed Gaussian random variables, where

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

What is the distribution of $X + Y$?

- (d) 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 $\mathbb{E}[|H - T|]$?

2 Linear Algebra

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

3 Calculus and optimization

- (a) Let $\mathbf{A} \in \mathbb{R}^{d \times D}$ be a real matrix, and $\vec{b} \in \mathbb{R}^d$ be a real vector. Define the function

$$f : \mathbb{R}^D \rightarrow \mathbb{R}$$
$$\vec{x} \mapsto \|\mathbf{A}\vec{x} - \vec{b}\|^2 + \|\vec{x}\|^2.$$

- (i) What is $\nabla f(\vec{x})$?
 - (ii) What value of \vec{x} minimizes f ? That is, find $\arg \min_{\vec{x}} f(\vec{x})$.
(Hint: Compute the stationary points of f .)
- (b) Let $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined as
- $$g(x) := x^\top A x - b^\top x + c,$$
- 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 Programming practice

- (a) Download the MATLAB data file `hw0data.mat`. Write a script that does the following.

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

- (i) Load the data in `hw0data.mat`, which 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 $\mathbf{M}^\top \mathbf{M}$.
- (b) 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 \mathbf{L} on random vectors and relate it to eigenvectors and eigenvalues.
- Write a script that does the following.

1. Create the 2×2 matrix \mathbf{L} (as defined above).
2. 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 $\mathcal{N}(0, 1)$ 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.
3. For each vector $\vec{r} \in R$, compute how the matrix \mathbf{L} “distorts” \vec{r} . That is, compute $\tilde{\mathbf{r}} := \mathbf{L}\vec{r}$.
4. Compute the eigenvalues of \mathbf{L} .
5. For each distorted vector $\tilde{\mathbf{r}}$, compute the length $\|\tilde{\mathbf{r}}\|$.
 - (i) What are the maximum and the minimum eigenvalues λ_{\max} and λ_{\min} ?
 - (ii) Create a histogram of values of $\|\tilde{\mathbf{r}}\|$ (use 50 bins) and compare it to λ_{\max} and λ_{\min} .
 - (iii) What relationship can you infer between $\|\tilde{\mathbf{r}}\|$, λ_{\max} and λ_{\min} ?
 - (iv) Now, compute the eigenvectors of \mathbf{L} . Let \vec{v}_{\max} denote the eigenvector corresponding to the maximum eigenvalue λ_{\max} .
Create a two-dimensional plot of all the distorted vectors $\tilde{\mathbf{r}}$ (in black color) and the eigenvector $\mathbf{L}\vec{v}_{\max}$ (in red color). (make sure that the x - and the y -axes are displayed at the same scale).
What can you infer about the \vec{v}_{\max} from studying this plot?