COMS 4771 HW0

Due: Sat Jan 25, 2020 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.

Notation

- $\Pr[\cdot]$ denotes the probability (of an event).
- $\mathbb{E}[\cdot]$ denotes the expected value (of a random variable).
- var$[\cdot]$ denotes the variance (of a random variable).
- 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 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
X=1	0.1	0.2
X=2	0.2	0.1
X=3	0.3	0.1

(i) What is the marginal distribution of X?

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

(iii) What is the variance of Y?

(iv) Let $f : x \mapsto x^2$. What is $\mathbb{E}[f(X)|Y = 1]$?

(v) Continuing form part (iv), what is $\mathbb{E}[f(X)|Y]$?

(b) Let $\theta > 0$ be an arbitrary positive number, and consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{Z}(e^{-\theta x} + e^{-2\theta x}) & \text{if } x \geq 0 \end{cases}$$

where $Z > 0$ is some positive constant that depends only on θ. Let X be a continuous random variable with probability density proportional to f.

(i) For what value of Z is f a probability density?

(ii) What is the mean of X?

(iii) What is the variance of X?

(In each case, give your answer in terms of θ.)

2 Linear Algebra Basics

Let $A \in \mathcal{S}_n(\mathbb{R})$ be a real symmetric matrix of size $n \times n$. Let V be a subspace of $\mathbb{R}^n$, we say V is an invariant subspace of A if $\forall x \in V, Ax \in V$.

(i) Show that if V is an invariant subspace of A then $V^\perp$ is also an invariant subspace of A.

(ii) Show that: nullspace$(A) =$ span$(A)^\perp$.

(iii) Show that the eigenvalues of A are real.

(iv) Let $\lambda_{\min}$ be the smallest eigenvalue of A. Show for any $\epsilon > 0$, the matrix $\left(A+(\lambda_{\min}+\epsilon)I_{n\times n}\right)$ is full rank.

3 Linear Algebra and Optimization Basics

Let $g : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by $g(x) := \frac{1}{2}x^\mathsf{T} Ax - b^\mathsf{T} x + c$ where

$$A := \begin{bmatrix} \alpha & \alpha+2 \\ \alpha+2 & \alpha \end{bmatrix}, \quad b := \begin{bmatrix} \beta \\ 2\beta \end{bmatrix}, \quad \text{and } c := \gamma.$$

(i) Compute the determinant of A.

(ii) Show that when $\alpha > 0$, the eigenvalues of A are real and of opposite sign.

(iii) Show that for any β which makes the the vector b a unit length, we have $b^\mathsf{T} x \leq \|x\|$.

(iv) Show that when $\alpha = 3$, $\beta = 1$ and $\gamma = 9$, the function g has neither a minimum nor a maximum.

(v) Provide all settings for α, β and γ for which g neither has a minimum nor a maximum.
(e.g. $\gamma \geq 1, \alpha = [0, 5] \cup [6\gamma, +\infty), \beta \in \mathbb{R}$.)

4 Of M and Ms

The blue M&M was introduced in 1995[1]. Before then, the color mix in a bag of plain M&Ms was (30% Brown, 20% Yellow, 20% Red, 10% Green, 10% Orange, 10% Tan). Afterward it was (24% Blue , 20% Green, 16% Orange, 14% Yellow, 13% Red, 13% Brown).

A friend of yours has two bags of M&Ms, and he tells you that one is from 1994 and one from 1996. He won't tell you which is which, but he gives you one M&M from each bag. One is yellow and one is green. What is the probability that the yellow M&M came from the 1994 bag?

5 Calculus, Optimization and Linear Algebra

Let $J : \mathbb{R}^6 \to \mathbb{R}$ be the function defined by

$$J(x) := \|Ax - b\|_2^2 + \frac{1}{4}\|x\|_2^2, \qquad x \in \mathbb{R}^6,$$

where

- $A \in \mathbb{R}^{6\times 6}$ is the matrix defined as

$$A := \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

- $b := (1, \ldots, 1) \in \mathbb{R}^6$.

[1] https://en.wikipedia.org/wiki/M%26M%27s

(i) What is $\nabla J(x)$ (i.e., the gradient of J) evaluated at $x = (0, \ldots, 0)$?

(ii) Find a value of x that minimizes J (i.e., find a value in $\arg\min_{x \in \mathbb{R}^6} J(x)$). *Hint:* It suffices to find a critical point of f.

(iii) Write a program in a scientific programming language of your choice that implements the Richardson iteration to (approximately) find a minimizer of the function $g : \mathbb{R}^6 \to \mathbb{R}$, where

$$g(x) := \|Ax - b\|_2^2, \qquad x \in \mathbb{R}^6 .$$

The Richardson iteration starts with an initial vector $x^{(0)} \in \mathbb{R}^6$, and computes subsequent vectors using the following recursion:

$$x^{(k)} := x^{(k-1)} + \eta A^{\mathsf{T}}(b - Ax^{(k-1)}), \quad k = 1, 2, \ldots,$$

where $\eta > 0$ is a "step size" parameter. Run your program starting with $x^{(0)} := (0, \ldots, 0)$ and $\eta := 1/2$ for up to $k = 1000$ steps. Save the resulting 1000 vectors, as you will need them for the remaining parts of this problem. What are the vectors $x^{(100)}$ and $x^{(1000)}$ produced by your program?

(you must submit your code on Courseworks for full credit)

(iv) Continuing from part (iii), let $v := (0, 0, 0, 1, -1, 0)$ and $\tilde{x} := x^{(1000)} + v$. Compute $\|Ax^{(1000)} - b\|_2^2$ and $\|A\tilde{x} - b\|_2^2$. How do these values compare? Also compute $\|x^{(1000)}\|_2^2$ and $\|\tilde{x}\|_2^2$. How do these values compare?

(v) Run your program again starting with $x^{(0)} := (0, \ldots, 0)$, but this time with $\eta := 3/4$. Save the resulting 1000 vectors separately from those obtained in part (iii); we'll call them $y^{(1)}, \ldots, y^{(1000)}$. Make a plot with two curves. The first curve is $\|Ax^{(k)} - b\|_2^2$ as a function of k, for $1 \le k \le 100$. The second curve is $\|Ay^{(k)} - b\|_2^2$ as a function of k, for $1 \le k \le 100$. Use different line styles and different colors to distinguish the two curves. Add a legend to the plot that identifies each curve. Add an appropriate title to the plot, and add appropriate labels for the horizontal and vertical axes.