Important: This homework tests basic mathematical skills required for the course. Note that the order/points of questions do not always imply difficulty.

Submission: Please complete the homework and submit a PDF file (the only accepted file format). The *submission link* can be found here. Typed or scanned handwritten solutions are fine if they are easily readable. ***The deadline for this problem set is: Feb 12, 2020.***

Late Assignment Policy: Late problem sets will be levied a late penalty of 0.5% per hour (up to 72 hours). After 72 hours, no credit will be given.

Grading: The homework assignments in this course are self-graded. Once the solutions are available, please use them to assign points to your own **submitted** solution and submit the points. The *grading link* can be found here. A subset of the assignments will be chosen at random and double checked by the course staff. It is in your best interest to complete the assignments as they will prepare you for the exams. ***The deadline for submitting a grade for this problem set is: Feb 19, 2020.***

Total: 50 points.

Ziqi Tan U88387934

1 Basic Calculus [10 pts]

The following questions test your basic skills in computing the derivatives of univariate functions, as well as applying the concept of *convexity* to determine the properties of the functions.

(a) (3 pts) Find all extrema of the function $f(x) = ln(2 - x^2)$. For each extremum, state if it is a maximum or a minimum.

Domain: $2 - x^2 > 0 \Rightarrow -\sqrt{2} < x < \sqrt{2}$

$$\frac{df(x)}{dx} = \frac{-2x}{2-x^2} = 0 \Rightarrow x = 0$$

$$\begin{cases} \frac{df(x)}{dx} > 0, & \text{if } x < 0 \\ \frac{df(x)}{dx} < 0, & \text{if } x > 0 \end{cases}$$

$f(0) = \ln 2$ is the only extrema, which is a maximum.

(b) (3 pts) Show that $f(x) = \ln \frac{1}{1+e^{-x}}$ is concave.

Domain: $\mathbb{R}$

$$\frac{df(x)}{dx} = \frac{e^{-x}}{1+e^{-x}} \quad \frac{d^2f(x)}{dx^2} = \frac{-e^{-x}}{(1+e^{-x})^2} < 0$$

$\therefore f(x)$ is a concave function

(c) (4 pts) Show that $f(x) = e^{-x^2}$ is neither convex nor concave.

Domain: $\mathbb{R}$

$$\frac{df(x)}{dx} = -2xe^{-x^2} \quad \frac{d^2f(x)}{dx^2} = -2e^{-x^2} + (-2x)^2 e^{-x^2}$$
$$= (4x^2 - 2)e^{-x^2}$$

when $4x^2 - 2 < 0 \Leftrightarrow -\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2}$, $\frac{d^2f(x)}{dx^2} < 0$

when $4x^2 - 2 > 0 \Leftrightarrow x < -\frac{\sqrt{2}}{2}$ or $x > \frac{\sqrt{2}}{2}$, $\frac{d^2f(x)}{dx^2} > 0$

$\therefore f(x)$ is neither convex nor concave.

2 Continuous Random Variables [10 pts]

(a) (2 pts) Given a continuous random variable X with probability density function $f(X)$, what are the expressions for the mean and variance of this variable?

mean: $\mu(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx$, Let it be $E(X)$.

variance: $\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f_X(x)\,dx$

(b) (2 pts) Can the value of the probability density function (PDF) $f(X)$ exceed 1? Why or why not?

Yes.

From definition, PDF's integral over the entire sample space is equal to 1, i.e. $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$

(c) (2 pts) Consider a random variable X that follows the *uniform distribution* between a and b, i.e. its PDF is equal to a constant c on this interval, and 0 otherwise. Derive c in terms of a and b.

$$f_X(x) = \begin{cases} c, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

$\because \int_{-\infty}^{+\infty} f_X(x)\,dx = (b-a)c = 1$

$\therefore c = \dfrac{1}{b-a}$

(d) (2 pts) Derive the expected value of X in terms of a and b. Show all your steps.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$
$$= \int_a^b x \frac{1}{b-a}\, dx$$
$$= \frac{1}{b-a} \int_a^b x\, dx$$
$$= \frac{1}{b-a} \cdot \frac{1}{2} \times (a+b)(b-a)$$
$$= \frac{a+b}{2}$$

(e) (2 pts) Derive the cumulative distribution function $F(X)$ on the interval $a \leq X \leq b$.

$$F(x) = \int_{-\infty}^{x} f_X(t)\, dt$$
$$= \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$\int_a^x \frac{1}{b-a}\, dt = \frac{t}{b-a}\Big|_a^x = \frac{x-a}{b-a} \quad (a \leq x \leq b)$$

3 Discrete Random Variables [10 pts]

(a) (2 pts) Two students taking a Machine Learning class became project partners. They are trying to decide what operating system to use for the project. Suppose each student has a laptop, which could be one of three types: Mac OS, Windows, or Linux. If the distribution of laptops among students follows the PDF shown below, what is the probability that the two teammates have **different** laptops?

Mac OS	0.6
Windows	0.3
Linux	0.1

Suppose: X is a random variable representing a laptop that a student has.

Event E_1 that two students have the same laptops has a probability of $P(E_1) = P(X = \text{Mac OS})^2 + P(X = \text{Windows})^2 + P(X = \text{Linux})^2$
$= 0.6^2 + 0.3^2 + 0.1^2 = 0.46$

Event E_2 that two students have different laptops has a probability of $P(E_2) = 1 - P(E_1) = 1 - 0.46 = 0.54$

Suppose we have three discrete random variables x, y and z that take values 0 or 1 according to the distribution below.

$x = 0$

	$z = 0$	$z = 1$
$y = 0$	0	$\frac{1}{12}$
$y = 1$	$\frac{1}{4}$	$\frac{1}{4}$

$x = 1$

	$z = 0$	$z = 1$
$y = 0$	$\frac{1}{12}$	$\frac{1}{12}$
$y = 1$	0	$\frac{1}{4}$

(b) (2 pts) Find the joint distribution of y and z

	$z=0$	$z=1$
$y=0$	$\frac{1}{12} = \frac{1}{12} + 0$	$\frac{1}{6} = \frac{1}{12} + \frac{1}{12}$
$y=1$	$\frac{1}{4} = \frac{1}{4} + 0$	$\frac{1}{2} = \frac{1}{4} + \frac{1}{4}$

(c) (2 pts) Find the marginal distributions of y and z

$P(Y=0) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$ $P(Y=1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

$P(Z=0) = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$ $P(Z=1) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$

(d) (2 pts) Find the conditional distribution of x given that $y = 0$.

$P(X=0) = 0 + \frac{1}{12} + \frac{1}{4} + \frac{1}{4} = \frac{7}{12}$

$P(X=1) = \frac{1}{12} + \frac{1}{12} + 0 + \frac{1}{4} = \frac{5}{12}$

$$P(X=0 \mid Y=0) = \frac{P(X=0 \cap Y=0)}{P(Y=0)} = \frac{0+\frac{1}{12}}{\frac{1}{4}} = \frac{1}{3}$$

$$P(X=1 \mid Y=0) = \frac{P(X=1 \cap Y=0)}{P(Y=0)} = \frac{\frac{1}{12}+\frac{1}{12}}{\frac{1}{4}} = \frac{2}{3}$$

(e) (2 pts) Are y and z independent? Explain.

"y" and "z" are independent if and only if

$P(Y \cap Z) = P(Y)P(Z)$

Refer to ~~the~~ joint distribution of y and z.

	$Z=0$	$Z=1$	$P(Y)$
$Y=0$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$
$Y=1$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
$P(Z)$	$\frac{1}{3}$	$\frac{2}{3}$	

$$P(Y=0 \cap Z=0) = \frac{1}{12} = P(Y=0)P(Z=0) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$
$$P(Y=0 \cap Z=1) = \frac{1}{6} = P(Y=0)P(Z=1) = \frac{1}{4} \times \frac{2}{3} = \frac{1}{6}$$
$$P(Y=1 \cap Z=0) = \frac{1}{4} = P(Y=1)P(Z=0) = \frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$$
$$P(Y=1 \cap Z=1) = \frac{1}{2} = P(Y=1)P(Z=1) = \frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$$

$\therefore$ y and z are independent.

4 Basic Linear Algebra [10 pts]

(a) (3 pts) Let $\boldsymbol{A}$ be a 3x4 matrix, $\boldsymbol{B}$ be a 4x5 matrix, and $\boldsymbol{C}$ be a 4x4 matrix. Determine which of the following products are defined and find the size of those that are defined. Note, X^T refers to the transpose of X.

$\boldsymbol{AB}$ 3x5

$\boldsymbol{BA}$ X

$\boldsymbol{AC}$ 3x4

$\boldsymbol{CA^T}$ 4x3

$\boldsymbol{BC^T}$ X

$\boldsymbol{CB}$ 4x5

(b) (3 pts) Suppose we would like to predict the profits of "Sunny Coffee", a bakery chain with locations in three different cities. Given the price of flour x, price of sugar y and price of oil z, the profit can be modelled as a linear function of these variables. That is, for each of the locations $i = 1, ..., 3$, the profit is $p_i = a_i + b_i x + c_i y + d_i z$.

Write down the matrix-vector product that produces the 3-dimensional vector of profits for the three locations.

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$= \begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \\ z \end{bmatrix}$$

(c) (4 pts) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $\mathbb{R}^{\mathsf{D}\times\mathsf{D}}$ symmetric matrices. Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ have the exact same set of eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{\mathsf{D}}$ with the corresponding eigenvalues $\alpha_1, \alpha_2, \cdots, \alpha_{\mathsf{D}}$ for $\boldsymbol{A}$, and $\beta_1, \beta_2, \cdots, \beta_{\mathsf{D}}$ for $\boldsymbol{B}$. Write down the eigenvectors and their corresponding eigenvalues for the following matrices. (*Hint.* Represent $\boldsymbol{A}$, $\boldsymbol{B}$ using the eigenvectors, e.g., $\boldsymbol{A} = \sum_d \alpha_d \boldsymbol{u}_d \boldsymbol{u}_d^{\mathrm{T}}$.)

- $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$

$A u_d = \alpha_d u_d$

$B u_d = \beta_d u_d$

$(A+B) u_d = (\alpha_d + \beta_d) u_d$

$\therefore \; C u_d = (\alpha_d + \beta_d) u_d$

$\therefore$ C: eigenvalues $\alpha_d + \beta_d$
eigenvector u_d
where $d = 1, 2, \cdots, D$

- $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$

$(A - B) u_d = (\alpha_d - \beta_d) u_d$

$\therefore \; D u_d = (\alpha_d - \beta_d) u_d$

$\therefore$ D: eigenvalues $\alpha_d - \beta_d$
eigenvectors u_d where $d = 1, 2, \cdots, D$

- $\boldsymbol{E} = \boldsymbol{A}\boldsymbol{B}$

$$\begin{aligned} E u_d &= A B u_d \\ &= A \beta_d u_d \\ &= \beta_d A u_d \\ &= \alpha_d \beta_d u_d \end{aligned}$$

$\therefore$ eigenvalues: $\alpha_d \beta_d$
eigenvectors: u_d
where $d = 1, 2, \cdots, D$

- $\boldsymbol{F} = \boldsymbol{A}^{-1}\boldsymbol{B}$ (assume $\boldsymbol{A}$ is invertible)

$$\begin{aligned} A^{-1} A B u_d &= A^{-1} A \beta_d u_d \\ &= A^{-1} \alpha_d \beta_d u_d \\ &= A^{-1} B \alpha_d \\ &= F \alpha_d \\ &= B u_d \end{aligned}$$

$\therefore F \alpha_d u_d = \beta_d u_d$

$\therefore F u_d = \dfrac{\beta_d}{\alpha_d} u_d$

$\therefore$ eigenvalues: $\dfrac{\beta_d}{\alpha_d}$
eigenvectors: u_d where $d = 1, 2, \cdots, D$

5 Vector Calculus [10 pts]

Consider the quadratic function $f(x) = x^T A x$ where x is a column vector and A is an nxn constant matrix.

$1 \times n \quad n \times n \quad n \times 1$ $\quad n \times n$

(a) (1 pts) Express $f(x)$ as a sum of terms (hint: use Σ).

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

(b) (4 pts) Compute the partial derivative of the function with respect to the kth element of x, i.e. $\dfrac{\partial f(x)}{\partial x_k}$, using the expression from (a). Express your answer as a sum of terms.

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i=1}^{n} A_{ik} x_i + \sum_{i=1}^{n} A_{ki} x_i = \sum_{i=1}^{n} (A_{ik} + A_{ki}) x_i$$

See the sketch below

Sketch:

Assume A is a 3 by 3 matrix.

$$f(x) = x^T A x$$

$$= [x_1 \; x_2 \; x_3] \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}x_1 + A_{21}x_2 + A_{31}x_3 \\ A_{12}x_1 + A_{22}x_2 + A_{32}x_3 \\ A_{13}x_1 + A_{23}x_2 + A_{33}x_3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \left[\sum_{j=1}^{3} A_{j1}x_j \;\; \sum_{j=1}^{3} A_{j2}x_j \;\; \sum_{j=1}^{3} A_{j3}x_j \right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \left(\sum_{j=1}^{3} A_{j1}x_j\right)x_1 + \left(\sum_{j=1}^{3} A_{j2}x_j\right)x_2 + \left(\sum_{j=1}^{3} A_{j3}x_j\right)x_3$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{3} A_{ji}x_j x_i$$

in general if A is a $n \times n$ matrix

$$f(x) = x^T A x$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} x_i x_j$$

$$\frac{\partial f(x)}{\partial x_1} = \underline{2A_{11}x_1 + A_{21}x_2 + A_{31}x_3} + A_{12}x_2 + A_{13}x_3$$

$$= \sum_{j=1}^{3} A_{j1}x_j + A_{11}x_1 + A_{12}x_2 + A_{13}x_3$$

$$= \sum_{j=1}^{3} A_{j1}x_j + \sum_{j=1}^{3} A_{1j}x_j$$

$$= \begin{bmatrix} 2A_{11} \\ A_{21}+A_{12} \\ A_{31}+A_{13} \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

in general $\frac{\partial f(x)}{\partial x_k} = \sum_{j=1}^{n} A_{jk}x_j + \sum_{j=1}^{3} A_{kj}x_j$

(c) (2 pts) Now write down the gradient vector $\nabla_x f(x)$ in matrix/vector notation, using the answer from (b). What is its dimension and meaning?

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} (A_{i1} + A_{1i}) x_i \\ \sum_{i=1}^{n} (A_{i2} + A_{2i}) x_i \\ \vdots \\ \sum_{i=1}^{n} (A_{in} + A_{ni}) x_i \end{bmatrix}$$

dimension : $n \times 1$,
Each element is a partial derivative of $f(x)$.

(d) (3 pts) Compute the second derivative of $f(x)$, $\nabla_x^2 f(x)$, in matrix form.

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

$$= \begin{bmatrix} 2A_{11} & A_{12} + A_{21} & \cdots & A_{1n} + A_{n1} \\ A_{21} + A_{12} & A_{22} & \cdots & A_{2n} + A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} + A_{1n} & A_{n2} + A_{2n} & \cdots & 2A_{nn} \end{bmatrix}$$

$$= A + A^T$$