Important: This homework tests basic mathematical skills required for the course. Use it as a self-test of your background knowledge: if you are having to look things up a lot, or if more than 20% of concepts are unfamiliar to you, you will struggle in the course. Note that the order/points of questions do not always imply difficulty.

Grading: The homework assignments in this course are self-graded. Please complete the homework and submit a PDF file (**the only accepted file format**). Typed or scanned handwritten solutions are fine. Once the solutions are available, please use them to assign points to your own **submitted** solution and submit the scores. Refer to the main course website for due dates and submission links. 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 individually as they will prepare you for the exams.

1 Basic Calculus

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) (2 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.

$$f'(x) = \frac{1}{(2-x^2)}(-2x) = 0 \quad (x \in (-\sqrt{2}, \sqrt{2}))$$

$$x_1 = 0$$

$\because$ when $x \in [0, \sqrt{2})$, $f'(x) < 0$; when $x \in (-\sqrt{2}, 0)$, $f'(x) > 0$

$\therefore f(x_1) = f(0)$ is the maximum extrema

$f(0) = \ln 2$

(b) (2 pts) Show that $f(x) = \ln \frac{1}{1+e^{-x}}$ is concave.

$$f'(x) = (1+e^{-x}) \cdot \frac{-1}{(1+e^{-x})^2} \cdot (-e^{-x}) = 1 - \frac{1}{1+e^{x}} > 0$$

$$f''(x) = \frac{1}{(1+e^{-x})^2} \cdot (-e^{-x}) = \frac{-e^{-x}}{(1+e^{-x})^2} < 0$$

$\because$ the second derivative $f''(x) < 0$

$\therefore f(x)$ is concave

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

$$f'(x) = e^{-x^2} \cdot (-2x) = -2xe^{-x^2}$$

$$f''(x) = -2e^{-x^2} - 2x \cdot (e^{-x^2}) \cdot (-2x)$$

$$= -2e^{-x^2} + 4x^2 e^{-2x^2}$$

$\because$ when $f''(x) = 0$, $x = \pm \frac{1}{\sqrt{2}}$

$$\therefore f''(x) \begin{cases} > 0 & x \in (-\infty, -\frac{1}{\sqrt{2}}) \\ < 0 & x \in (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \\ > 0 & x \in (\frac{1}{\sqrt{2}}, +\infty) \end{cases}$$

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

2 Continuous Random Variables

(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/expectation of X : $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$

variance of X : $D(x) = E(x^2) - E^2(x)$

(b) (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.

$$f(x) = \begin{cases} c & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$$c \cdot (b-a) = 1$$

$$c = \frac{1}{b-a}$$

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

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

$$= \frac{1}{b-a} \left(\frac{1}{2}x^2\right)\Big|_a^b = \frac{b+a}{2}$$

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

$$F(x) = \int_a^x f(x) dx = \frac{1}{b-a} x \Big|_a^x = \frac{x-a}{b-a} \quad (a \leq x \leq b)$$

3 Discrete Random Variables

(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

① assume student A, B use the same OS

$P = 0.6^2 + 0.3^2 + 0.1^2 = 0.46$

then $1-P = 0.54$ is the probability of A, B using different OS

② ~~probabil~~ probability of having differe OS/laptops is

$P = 0.6\cdot(1-0.6) + 0.3\cdot(1-0.3) + 0.1(1-0.1) = 0.54$

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

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

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

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

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

$$P(x \mid y=0) = \begin{cases} P(x=0 \mid y=0) = \dfrac{P(x=0, y=0)}{P(y=0)} = \dfrac{\frac{1}{12}}{\frac{1}{4}} = \dfrac{1}{3} \\ P(x=1 \mid y=0) = \dfrac{P(x=1, y=0)}{P(y=0)} = \dfrac{\frac{1}{6}}{\frac{1}{4}} = \dfrac{2}{3} \end{cases}$$

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

Yes. because the joint probability of y, z is equal to the multiplication of ~~th~~ individual's marginal distributions.

~~$P(X=x_1, Y=$~~

$$P(Y=y_i, Z=z_j) = P(y_i) \cdot P(z_j)$$

4 Basic Linear Algebra

(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.

$A_{3\times4}$
$B_{4\times5}$
$C_{4\times4}$

$\boldsymbol{AB}$ ✓ 3×5 $\boldsymbol{BA}$ ✗

$\boldsymbol{AC}$ ✓ 3×4 $\boldsymbol{CA^T}$ ✓ 4×3

$\boldsymbol{BC^T}$ ✗ $\boldsymbol{CB}$ ✓ 4×5

(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 $i = 1, \ldots, 3$, 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.

$$Ax = b, \text{ where } A = \begin{pmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{pmatrix} \quad x = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$b = \begin{pmatrix} p_1 - a_1 \\ p_2 - a_2 \\ p_3 - a_3 \end{pmatrix}$$

(c) (4 pts) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $\mathbb{R}^{D\times D}$ symmetric matrices. Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ have the exact same set of eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_D$ with the corresponding eigenvalues $\alpha_1, \alpha_2, \cdots, \alpha_D$ for $\boldsymbol{A}$, and $\beta_1, \beta_2, \cdots, \beta_D$ 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., $\boldsymbol{A} = \sum_d \alpha_d \boldsymbol{u}_d \boldsymbol{u}_d^T$.)

Define: $P = (\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4 \cdots \vec{u}_D)$

$\lambda_\alpha =$ diagonal $(\alpha_1 \cdots \alpha_D)$

$\lambda_\beta =$ diagonal $(\beta_1 \cdots \beta_D)$

then $A = P\lambda_\alpha P^{-1}$, $B = P\vec{\lambda}_\beta P^{-1}$

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

$$C = A+B = P\vec{\lambda}_\alpha P^{-1} + P\vec{\lambda}_\beta P^{-1}$$
$$= P(\vec{\lambda}_\alpha + \vec{\lambda}_\beta)P^{-1}$$

eigenvectors: $\vec{u}_1 \cdots \vec{u}_D$

eigenvalues: $\alpha_1+\beta_1, \alpha_2+\beta_2 \cdots \alpha_D+\beta_D$

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

$$D = A-B = P(\vec{\lambda}_\alpha - \vec{\lambda}_\beta)P^{-1}$$

eigenvalues: $\alpha_1-\beta_1, \cdots \alpha_D-\beta_D$

eigenvectors: $\vec{u}_1 \cdots \vec{u}_D$

- $\boldsymbol{E} = \boldsymbol{AB}$

$$E = A\cdot B = P\cdot\vec{\lambda}_\alpha\cdot\vec{\lambda}_\beta\cdot P^{-1}$$

eigenvalues: $\alpha_1\beta_1, \alpha_2\beta_2 \cdots \alpha_n\beta_n$

eigenvectors: $\vec{u}_1 \cdots \vec{u}_D$

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

$$F = (P\vec{\lambda}_\alpha P^{-1})^{-1} P\vec{\lambda}_\beta P^{-1}$$
$$= P\vec{\lambda}_\alpha^{-1} \cancel{P^{-1}} \cdot \cancel{P}\vec{\lambda}_\beta P^{-1}$$
$$= P\vec{\lambda}_\alpha^{-1}\vec{\lambda}_\beta P^{-1}$$

eigenvalues: $\frac{\beta_1}{\alpha_1} \cdots \frac{\beta_D}{\alpha_D}$

eigenvectors: $\vec{u}_1 \cdots \vec{u}_D$