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.

$f'(x) = \frac{2x}{x^2-2}$ $\quad f''(x) = \frac{-2(x^2+2)}{(x^2-2)^2}$

$f''(0) < 0$, so $f(x)$ at $x = 0$ is a maximum

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

$f'(x) = \frac{1}{e^x+1}$

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

$f''(x) < 0$, always negative.

$\therefore f(x)$ is concave.

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

$f'(x) = -2xe^{-x^2}$ $\quad f''(x) = (4x^2-2)e^{-x^2}$

$f''(0) = -2 < 0$ $\quad f''(1) = \frac{2}{e} > 0$

$\therefore f(x)$ 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?

$$u(x) = \int_{-\infty}^{\infty} x f(x) dx$$ (mean)

$$\sigma = E((X - E(x))^2)$$ (variance)

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

Yes, but the area below the curve has to be smaller than 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.

$$\int_a^b c \, dx = 1$$, PDF integrate to 1

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

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

$$\int_a^b \frac{x}{b-a} dx$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]\Big|_a^b$$

$$= \frac{1}{2} \cdot \frac{b^2-a^2}{b-a}$$

$$= \frac{a+b}{2}$$

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

$$F(x) = \int_a^x f(x)\,dx$$

$$= \int_a^x \frac{1}{b-a} dx$$

$$= \frac{x}{b-a} - \frac{a}{b-a}$$

$$= \frac{x-a}{b-a} \quad \text{for } a \le x \le b$$

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

$P_{same} = (0.6)^2 + (0.3)^2 + (0.1)^2 = 0.46$

$P_{diff} = 1 - 0.46 = 0.54$

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

By sum rule

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

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

By sum rule

	$z=0$	$z=1$	$z=$ total
$y=0$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$
$y=1$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
y total	$\frac{1}{3}$	$\frac{2}{3}$	1

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

$y=0$	$x=0$	$x=1$
	$\frac{1}{3}$	$\frac{2}{3}$

$p(x, y=0) = p(x|y=0)p(y=0)$,

$p(x|y=0) = p(x, y=0)/p(y=0)$

$p(x=0|y=0) = \frac{1}{3}$

$p(x=1|y=0) = \frac{2}{3}$.

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

Yes. $p(y,z) = p(y)p(z)$. for all y, z.

they are independent probabilities.

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

$\boldsymbol{BA}$ undefined

$\boldsymbol{AC}$ 3x4

$\boldsymbol{CA^T}$ 4x3

$\boldsymbol{BC^T}$ undefined

$\boldsymbol{CB}$ 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.

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

(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., $\boldsymbol{A} = \sum_d \alpha_d \boldsymbol{u}_d \boldsymbol{u}_d^{\mathrm{T}}$.)

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

$$C = \Sigma_d \alpha_d v_d v_d^T + \Sigma_d \beta_d v_d v_d^T = \Sigma_d (\alpha_d + \beta_d) v_d v_d^T$$

C's eigenvectors $= v_d$

C's eigenvalues $= (\alpha_d + \beta_d)$

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

$$C = \Sigma_d (\alpha_d - \beta_d) v_d v_d^T$$

C's eigenvectors $= v_d$

C's eigenvalues $= (\alpha_d - \beta_d)$

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

$$E = (\Sigma_i \alpha_i v_i v_i^T)(\Sigma_j \beta_j v_j v_j^T) = \Sigma_{i,j} \alpha_i \beta_j v_i v_i^T v_j v_j^T$$

E's eigenvectors $= v_d$

E's eigenvalues $= (\alpha_d \beta_d)$

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

$$E = \Sigma_{i,j} \frac{\beta_j}{\alpha_i} v_i v_i^T v_j v_j^T$$

E's eigenvectors $= v_d$

E's eigenvalues $= \dfrac{\beta_d}{\alpha_d}$

5 Vector Calculus [10 pts]

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

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

$$f(x) = \sum_{i=1}^{n} (a_i \cdot x)\, x_i = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j} x_i x_j$$

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

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i \neq k} \frac{\partial}{\partial x_k}\left(\sum_{j=1}^{n} a_{ij} x_i x_j\right) + \frac{\partial}{\partial x_k}\left(\sum_{j=1}^{n} a_{kj} x_k x_j\right)$$

$$= \sum_{i=1}^{n} a_{ik} x_i + \sum_{j=1}^{n} a_{jk} x_j$$

(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?

$$\nabla_x f(x) = (A + A^T)x$$

Dimension : $N \times 1$

Meaning : the gradient vector is towards largest direction and the magnitude is the rate of increase.

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

$$\nabla_x^2 f(x) = \frac{\partial((A+A^T)x)}{\partial x} = A + A^T$$