COMP9417 - Machine Learning
Homework 0: Revision

Introduction The goal of this homework is to review some important mathematical concepts that are used regularly in machine learning, and which are assumed knowledge for the course. If you find yourself struggling significantly with any aspects of this homework, please reach out to course staff so that we can better help you prepare for the course. Please also note that we have posted some helpful resources under the Week 0 tab on Moodle which may be of use to you for this homework.

What to Submit

- A single PDF file which contains solutions to each question. For each question, provide your solution in the form of text and requested plots. For any question in which you use code, provide a copy of your code at the bottom of the relevant section.
- You are free to format your work in any way you think is appropriate. This can include using LaTeX, or taking pictures of handwritten work, or writing your solutions up using a tablet. Please ensure that your work is neat, and start each question on a new page.

When and Where to Submit

- Due date: Friday **June 4th**, 2021 by **5:00pm**.
- For this homework, we will **not** accept late submissions.
- Submissions must be through Moodle - email submissions will be ignored.

Question 1. (Calculus Review)

(a) Consider the function

$$f(x, y) = a_1x^2y^2 + a_4xy + a_5x + a_7$$

compute all first and second order derivatives of f with respect to x and y.

(b) Consider the function

$$f(x, y) = a_1x^2y^2 + a_2x^2y + a_3xy^2 + a_4xy + a_5x + a_6y + a_7$$

compute all first and second order derivatives of f with respect to x and y.

(c) Consider the logistic sigmoid:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

show that $\sigma'(x) = \frac{\partial \sigma}{\partial x} = \sigma(x)(1 - \sigma(x))$

(d) Consider the following functions:

- $y_1 = 4x^2 - 3x + 3$
- $y_2 = 3x^4 - 2x^3$
- $y_3 = 4x + \sqrt{1-x}$
- $y_4 = x + x^{-1}$

Using the second derivative test, find all local maximum and minimum points.

Question 2. (Probability Review)

(a) A manufacturing company has two retail outlets. It is known that 20% of potential customers buy products from Outlet I alone, 10% buy from both I and II, and 40% buy from neither. Let A denote the event that a potential customer, randomly chosen, buys from outel I, and B the event that the customer buys from outlet II. Compute the following probabilities:

20%

$$P(A), \quad P(B), \quad P(A \cup B), \quad P(\bar{A}\bar{B})$$

1- 0.2-0.1-0.4=0.3

(b) Let X, Y be two discrete random variables, with joint probability mass function $P(X = x, Y = y)$ displayed in the table below:

		$y=1$	$y=2$	$y=3$
x	1	1/6	1/12	1/12
	2	1/6	0	1/6
	3	0	r	0

Compute the following quantities:

(i) r
(ii) $P(X = 2, Y = 3)$
(iii) $P(X = 3)$ and $P(X = 3|Y = 2)$
(iv) $\mathbb{E}[X]$, $\mathbb{E}[Y]$ and $\mathbb{E}[XY]$
(v) $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2]$
(vi) $\text{Cov}(X, Y)$
(vii) $\text{Var}(X)$ and $\text{Var}(Y)$
(viii) $\text{Corr}(X, Y)$
(ix) $\mathbb{E}[X + Y]$, $\mathbb{E}[X + Y^2]$, $\text{Var}(X + Y)$ and $\text{Var}(X + Y^2)$.

Question 3. (Linear Algebra Review)

(a) Write down the dimensions of the following objects:

$$A = \begin{bmatrix} 1 & 3 & 1 & 0 & 2 \\ 1 & 1 & 4 & 1 & 2 \\ 1 & 1 & 1 & 5 & 2 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \\ 3 \\ 2 \end{bmatrix}, \qquad A^T$$

(b) Consider the following objects:

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 2 & 1 \\ 6 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 \\ 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 7 & 3 & 3 \\ 2 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 4 & 2 \\ 4 & 6 \\ 1 & 3 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

Compute the following:

(i) AB and BA

(ii) AC and CA

(iii) AD and DA

(iv) DC and CD and D^TC

(v) Bu and uB

(vi) Au

(vii) Av and vA

(viii) $Av + Bv$

(c) Consider the following objects:

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 2 & 1 \\ 6 & 4 & 3 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}, \quad w = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}.$$

Compute the following:

(i) $\|u\|_1, \|u\|_2, \|u\|_2^2, \|u\|_\infty$

(ii) $\|v\|_1, \|v\|_2, \|v\|_2^2, \|v\|_\infty$

(iii) $\|v + w\|_1, \|v + w\|_2, \|v + w\|_\infty$

(iv) $\|Av\|_2, \|A(v - w)\|_\infty$

(d) Consider the following vectors in $\mathbb{R}^2$

$$u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad w = \begin{bmatrix} -1 \\ 1/2 \end{bmatrix}$$

Compute the dot products between all pairs of vectors. Note that the dot product may be written using the following equivalent forms:

$$\langle x, y \rangle = x \cdot y = x^T y.$$

Then compute the angle between the vectors and plot.

(e) Dot products are extremely important in machine learning, explain what it means for a dot product to be zero, positive or negative.

(f) Consider the 2×2 matrix:

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 1 \end{bmatrix}$$

Compute the inverse of A.

(g) Consider the 2×2 matrix

$$A = \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix}$$

Compute its inverse A^{-1}.

(h) Let X be a matrix (of any dimension), show that X^TX is always symmetric.

Question 1.

(a) $f(x,y) = a_1x^2y^2 + a_4xy + a_5x + a_7$

$$\frac{\partial}{\partial x} f(x,y) = 2a_1xy^2 + a_4y + a_5$$

$$\frac{\partial^2}{\partial x^2} f(x,y) = 2a_1y^2$$

$$\frac{\partial}{\partial y} f(x,y) = 2a_1x^2y + a_4x$$

$$\frac{\partial^2}{\partial y^2} f(x,y) = 2a_1x^2$$

(b) $f(x,y) = a_1x^2y^2 + a_2x^2y + a_3xy^2 + a_4xy + a_5x + a_6y + a_7$

$$\frac{\partial}{\partial x} f(x,y) = 2a_1xy^2 + 2a_2xy + a_3y^2 + a_4y + a_5$$

$$\frac{\partial^2}{\partial x^2} f(x,y) = 2a_1y^2 + 2a_2y$$

$$\frac{\partial}{\partial y} f(x,y) = 2a_1x^2y + a_2x^2 + 2a_3xy + a_4x + a_6$$

$$\frac{\partial^2}{\partial y^2} f(x,y) = 2a_1x^2 + 2a_3x$$

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

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

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

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

$$= \sigma(x) - \sigma^2(x)$$

$$= \sigma(x)(1-\sigma(x))$$

(d) $y_1 = 4x^2 - 3x + 3$

$$\begin{cases} y_1' = 8x - 3 \\ y_1'' = 8 \end{cases}$$

Hence, $x = \frac{3}{8}$ when $y_1' = 0$, and $y_1'' = 8 > 0$

Hence the local minimum points is $\left(\frac{3}{8}, \frac{39}{16}\right)$, and there is no local maximum points.

$$y_2 = 3x^4 - 2x^3$$

$$\begin{cases} y_2' = 12x^3 - 6x^2 \\ y_2'' = 36x^2 - 12x \end{cases}$$

Hence, $x = 0$ or $\frac{1}{2}$ when $y_2' = 0$.

When $x = 0$, $y_2'' = 0$, hence there is no local minimum or maximum points at $x = 0$.

When $x = \frac{1}{2}$, $y_2'' = 3 > 0$, hence the local minimum points is $\left(\frac{1}{2}, -\frac{1}{16}\right)$, and there is no local maximum points.

$$y_3 = 4x + \sqrt{1 - x}$$

$$\begin{cases} y_3' = 4 - \frac{1}{2} \cdot (1-x)^{-\frac{1}{2}} \\ y_3'' = -\frac{1}{4}(1-x)^{-\frac{3}{2}} \end{cases}$$

Hence, $x = \frac{63}{64}$ when $y_3' = 0$.

When $x = \frac{63}{64}$, $y_2'' < 0$, hence the local maximum points is $\left(\frac{63}{64}, 4.062\right)$ and there is no local minimum points.

$$y_4 = x + x^{-1}$$

$$\begin{cases} y_4' = 1 - x^{-2} \\ y_4'' = 2x^{-3} \end{cases}$$

Hence $x = \pm 1$ when $y_4' = 0$

When $x = 1$, $y_4'' > 0$, hence the local minimum points is $(1, 2)$,

When $x = -1$, $y_4'' < 0$, hence the local maximum points is $(-1, -2)$,

Question 2

(a) $P(A) = 20\%$

$P(B) = 1 - 20\% - 10\% - 40\% = 30\%$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 20\% + 30\% - 10\% = 40\%$

$P(\bar{A}\bar{B}) = 1 - P(A \cup B) = 1 - 40\% = 60\%$

(b) (i) $r = 1 - \frac{1}{6} - \frac{1}{12} - \frac{1}{12} - \frac{1}{6} - \frac{1}{6} = \frac{1}{3}$

(ii) $P(X=2, Y=3) = \frac{1/6}{1/12 + 1/6 + 0} = \frac{2}{3}$

(iii) $P(X=3) = r = \frac{1}{3}$, $P(X=3 \mid Y=2) = \frac{\frac{1}{3}}{\frac{1}{12}+\frac{1}{3}} = \frac{4}{5}$

(iv) $E(X) = (\frac{1}{6}+\frac{1}{12}+\frac{1}{12}) \times 1 + (\frac{1}{6}+\frac{1}{6}) \times 2 + \frac{1}{3} \times 3 = 2$

$E(Y) = (\frac{1}{6}+\frac{1}{6}) \times 1 + (\frac{1}{12}+\frac{1}{3}) \times 2 + (\frac{1}{12}+\frac{1}{6}) \times 3 = \frac{23}{12}$

$E(XY) = 1 \times 1 \times \frac{1}{6} + 1 \times 2 \times \frac{1}{12} + 1 \times 3 \times \frac{1}{12} +$

$2 \times 1 \times \frac{1}{6} + 2 \times 2 \times 0 + 2 \times 3 \times \frac{1}{6} +$

$3 \times 1 \times 0 + 3 \times 2 \times \frac{1}{3} + 3 \times 3 \times 0 = \frac{47}{12}$

(v) $E(X^2) = (\frac{1}{6}+\frac{1}{12}+\frac{1}{12}) \times 1^2 + (\frac{1}{6}+\frac{1}{6}) \times 2^2 + \frac{1}{3} \times 3^2 = \frac{14}{3}$

$E(Y^2) = (\frac{1}{6}+\frac{1}{6}) \times 1^2 + (\frac{1}{12}+\frac{1}{3}) \times 2^2 + (\frac{1}{12}+\frac{1}{6}) \times 3^2 = \frac{17}{4}$

(vi) $Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{47}{12} - 2 \times \frac{23}{12} = \frac{1}{12}$

(vii) $Var(X) = E(X^2) - E^2(X) = \frac{14}{3} - 2^2 = \frac{2}{3}$

$Var(Y) = E(Y^2) - E^2(Y) = \frac{17}{4} - (\frac{23}{12})^2 = \frac{83}{144}$

(viii) $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\frac{1}{12}}{\sqrt{\frac{2}{3}} \cdot \sqrt{\frac{83}{144}}} = 0.134$

(ix) $E(X+Y) = E(X) + E(Y) = 2 + \frac{23}{12} = \frac{47}{12}$

$E(X+Y^2) = E(X) + E(Y^2) = 2 + \frac{17}{4} = \frac{25}{4}$

$Var(X+Y) = Var(X) + Var(Y) - 2Cov(X,Y) = \frac{2}{3} + \frac{83}{144} - 2 \times \frac{1}{12} = \frac{155}{144}$

$Var(X+Y^2) = Var(X) + Var(Y^2) - 2Cov(X,Y^2) =$

Question 3

(a) The dimension of A is 3×5, b is 6×1, A^T is 5×3.

(b) (i) A is 3×3 and B is 2×2, hence both AB and BA cannot be computed.

(ii) $AC = \begin{bmatrix} 21 & 14 & 14 \\ 20 & 10 & 10 \\ 56 & 28 & 28 \end{bmatrix}$ $CA = \begin{bmatrix} 31 & 39 & 40 \\ 10 & 12 & 12 \\ 18 & 18 & 16 \end{bmatrix}$

(iii) $AD = \begin{bmatrix} 20 & 32 \\ 17 & 19 \\ 43 & 45 \end{bmatrix}$ DA cannot be computed.

(iv) DC cannot be computed. $CD = \begin{bmatrix} 43 & 41 \\ 13 & 13 \\ 18 & 22 \end{bmatrix}$ $D^TC = \begin{bmatrix} 38 & 18 & 18 \\ 32 & 18 & 18 \end{bmatrix}$

(v) $Bu = \begin{bmatrix} 14 \\ 4 \end{bmatrix}$, uB cannot be computed.

(vi) Au cannot be computed.

(vii) $Av = \begin{bmatrix} 18 \\ 13 \\ 31 \end{bmatrix}$, vA cannot be computed.

(viii) $Av + Bv$ cannot be computed due to Bv.

```python
import numpy as np

A = np.array([[1, 3, 4], [2, 2, 1], [6, 4, 3]])
B = np.array([[2, 4], [1, 1]])
C = np.array([[7, 3, 3], [2, 1, 1], [2, 2, 2]])
D = np.array([[4, 2], [4, 6], [1, 3]])
u = np.array([[1, 3]]).T
v = np.array([[2, 4, 1]]).T

result_2_1 = np.dot(A, C)
print(result_2_1)
result_2_2 = np.dot(C, A)
print(result_2_2)

result_3_1 = np.dot(A, D)
print(result_3_1)
# result_3_2 = np.dot(D, A)
# print(result_3_2)

# result_4_1 = np.dot(D, C)
# print(result_4_1)
result_4_2 = np.dot(C, D)
print(result_4_2)
result_4_3 = np.dot(D.T, C)
print(result_4_3)

result_5_1 = np.dot(B, u)
print(result_5_1)
# result_5_2 = np.dot(u, B)
# print(result_5_2)

# result_6_1 = np.dot(A, u)
# print(result_6_1)

result_7_1 = np.dot(A, v)
print(result_7_1)
# result_7_2 = np.dot(v, A)
# print(result_7_2)

result_8_1 = np.dot(A, v) + np.dot(B, v)
print(result_8_1)
```

(c) (i) $\|u\|_1 = |1| + |3| = 4$; $\|u\|_2 = \sqrt{1^2+3^2} = \sqrt{10}$; $\|u\|_2^2 = u^Tu = [10]$; $\|u\|_\infty = 3$

(ii) $\|v\|_1 = |2| + |4| + |1| = 7$; $\|v\|_2 = \sqrt{2^2+4^2+1^2} = \sqrt{21}$; $\|v\|_2^2 = v^Tv = [21]$; $\|v\|_\infty = 4$

(iii) $v+w = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$, $\|v+w\|_1 = |3| + |2| + |3| = 8$; $\|v+w\|_2 = \sqrt{3^2+2^2+3^2} = \sqrt{22}$

$\|v+w\|_\infty = 3$

(iv) $Av = \begin{bmatrix} 18 \\ 13 \\ 31 \end{bmatrix}$, $\|Av\|_2 = \sqrt{18^2+13^2+31^2} = 38.131$

$A(v-w) = \begin{bmatrix} 15 \\ 13 \\ 27 \end{bmatrix}$, $\|A(v-w)\| = 27$

(d) $u \cdot v = 1 \times 1 + 2 \times 1 = 3$, $u \cdot w = 1 \times (-1) + 2 \times \frac{1}{2} = 1$, $v \cdot w = 1 \times (-1) + 1 \times \frac{1}{2} = -\frac{1}{2}$

$\|u\|_2 = \sqrt{1+2^2} = \sqrt{5}$, $\|v\|_2 = \sqrt{2}$, $\|w\|_2 = \frac{\sqrt{5}}{2}$

$\cos(u,v) = \frac{3}{\sqrt{10}}$, $\cos(u,w) = \frac{2}{\sqrt{5}}$, $\cos(v,w) = \frac{-1}{\sqrt{10}}$

Hence, $18.558°$ for u, v, $26.62°$ for u, w, $108.42°$ for v, w

(e)

It can show the similarity of the items, The sign of the dot product determines the angle
0 means 90°, positive means less than 90°, negative means larger than 90°.

(f) $A^{-1} = \frac{1}{1\times 1 - 4\times 3}\begin{bmatrix} 1 & -3 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{11} & \frac{3}{3} \\ \frac{4}{11} & -\frac{1}{11} \end{bmatrix}$

(g) Due to $3\times 4 - 3\times 4 = 0$, there is no inverse A^{-1} for A.

(h) $(X^T X)^T = X^T \cdot (X^T)^T = X^T X$, hence $X^T X = (X^T X)^T$, and $X^T X$ is always symmetric.