

CS 556-B: Mathematical Foundations of Machine Learning Homework 2: Linear Algebra, Dimensionality Reduction, Probability & Differential Calculus (100 points)

Note: All solutions methods must be fully explained. If a problem requires you to submit your code, please ensure that your code (in the form of a jupyter notebook) is developed in a python3 environment and appropriately commented.

Eigenvalues & Eigenvectors, PCA & SVD (25 points)

- (5 points) Diagonalize the matrix $\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$ Please arrange the eigenvalues in descending order while constructing the diagonal matrix.

Answer: Using the equation $\det(A - \lambda I) = 0$

$$\det\left(\begin{bmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{bmatrix}\right) = 0 \implies (1-\lambda)(3-\lambda) - 2*4 = 0$$

$$3 - \lambda - 3\lambda + \lambda^2 - 8 = 0 \implies \lambda^2 - 4\lambda - 5 = 0 \implies (\lambda + 1)(\lambda - 5) = 0 \implies \lambda = 5, -1$$

$$\Lambda = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\text{For } \lambda = 5, \begin{bmatrix} 1-5 & 2 \\ 4 & 3-5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0} \implies \mathbf{v}_1 = 1 \text{ and } \mathbf{v}_2 = 2$$

$$\text{For } \lambda = -1, \begin{bmatrix} 1+1 & 2 \\ 4 & 3+1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0} \implies \mathbf{v}_1 = -1 \text{ and } \mathbf{v}_2 = 1$$

$$\implies V = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \implies V^{-1} = \begin{bmatrix} 1/3 & 1/3 \\ -2/3 & 1/3 \end{bmatrix}$$

$$A = V \Lambda V^{-1} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 \\ -2/3 & 1/3 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 1 \\ 10 & -1 \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 \\ -2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

- (5 points) Calculate the Singular Value Decomposition of matrix $A = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$.

Demonstrate each step in the computation and order the singular values in descending order in the Σ matrix.

$$\text{Answer: } \mathbf{A}^T \mathbf{A} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 13 & 12 \\ 12 & 13 \end{bmatrix}$$

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0 \implies \det\left(\begin{bmatrix} 13-\lambda & 12 \\ 12 & 13-\lambda \end{bmatrix}\right) = 0 \implies (13-\lambda)^2 - 144 = 0$$

$$(13-\lambda)^2 = 144 \implies (13-\lambda) = \pm\sqrt{144} \implies \lambda = 25, 1$$

$$\Sigma^2 = \begin{bmatrix} 25 & 0 \\ 0 & 1 \end{bmatrix} \implies \Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\text{For } \lambda = 25, \begin{bmatrix} 13-25 & 12 \\ 12 & 13-25 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0} \implies \mathbf{v}_1 = 1 \text{ and } \mathbf{v}_2 = 1$$

$$\text{For } \lambda = 1, \begin{bmatrix} 13-1 & 12 \\ 12 & 13-1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0} \implies \mathbf{v}_1 = 1 \text{ and } \mathbf{v}_2 = -1$$

$$V = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ } V \text{ is not orthogonal, so dividing it by the magnitude of its vectors.}$$

$$V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}, V^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

$$\Sigma^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$U = A V \Sigma^{-1} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

3. (8 points) Recommendation System (SVD)

	Rambo	Top-Gun	Harry Potter	Amelie	Casablanca
User 1	5.0	4.5	-	-	1.0
User 2	-	-	2.5	5.0	5.0
User 3	4.0	4.5	-	-	-
User 4	-	1.5	4.5	-	-
User 5	1.0	-	-	4.0	5.0
User 6	-	-	4.5	-	-

Table 1: Rating Matrix

Table 1 showcases user preferences of various movies. The rating scale is from 1 - 5 where 1 indicates a strong dislike of the movie while 5 indicates that the user has a strong liking for the movie. Some ratings have been masked ('-' symbol) and it is your task to re-create these ratings from the 'U','V','S' matrices provided (as numpy arrays stored on disk to be read and processed via. a python script).

- (1 point) What is U1's 'estimated' rating of the movie *Amelie*?
- (2 point) What is the *strength* of concept 1 in the SVD decomposition?
- (1 point) What is the average rating for movie Harry Potter across all users in the system (i.e., if actual ratings exist, use them if not use estimated ratings)?
- (1 point) What is the movie with the overall highest rating (i.e., if actual ratings exist, use them if not use estimated ratings)?
- (3 points) Please include your code, as a python3 jupyter notebook file named: '`<lastname>_<firstname>.svd.ipynb`'. All visualizations and answers should be printed in the notebook (in addition to including only the answers in the assignment for this question).

Answer: a) 1.4999999898595036

b) 13.266399703684902

c) 2.416666650522624

d) Casablanca

e) Check file *agrawal_anmol_svd.ipynb*

4. (7 points) Calculate the principle component analysis of the Iris dataset using the built-in method from Scikit-learn and showcase the following.

- (1 point) Include a 2D scatter plot (i.e., PCA with 2 components) where each point is colored by the class (i.e., category) to which it belongs.
- (1 point) What is the total percentage of variance captured by the first 2 components of PCA?
- (1 point) What is the *strength* of each of the two principal components?
- (1 point) What is the magnitude of each of the two principal components?
- (3 points) Please include your code, as a python3 jupyter notebook file named: '`<lastname>_<firstname>.pca.ipynb`'. All visualizations and answers should be printed in the notebook (in addition to including only the answers in the assignment for this question)

Answer: Check file *agrawal_anmol_pca.ipynb*

Probability (50 points)

5. (10 points) Jane picks apples. The weight of the apples picked by Jane are normally distributed with a mean of 180g and a standard deviation of 60g. Jane does not like apples that weigh less than or equal to 100g.
- (a) (5 points) What is the percentage of apples that Jane picks and does not like? **Hint:** Employ Z-scores and the z-table to calculate the required percentage.
- (b) (5 points) Let q_3 represent the minimum weight for the upper quartile (i.e., top 25%) of weight of apples liked by Jane. What is the probability that the weight of a randomly selected apple picked by Jane is greater than or equal to q_3 . **Hint:** Consider the probability that Jane *likes* an apple AND that the weight of the apple lies in the *upper quartile* (i.e., weight $\geq q_3$)

Answer: a) Jane doesn't like apples that weigh less than or equal to 100g. So, $Z(w \leq 100) = \frac{100-180}{60} = -1.33$
 $P(\text{Jane not liking an apple she picked}) = 0.0918$ (From the z-table). So, Jane doesn't like 9.18 percent of apples that she picked.

b) $P(\text{Jane likes an apple she picked}) = 1 - P(\text{Jane not liking an apple she picked}) = 1 - 0.0918 = 0.908$
 We have to find out the probability of an apple picked by Jane being in the upper quartile.
 So, $P(w \geq q_3) = 0.25 * P(\text{Jane liking the apple she picked}) = 0.25 * 0.908 = 0.227$

6. (10 points) A bag contains five types of balls colored red, blue, green, yellow, orange. There are 9 each of red, blue, green and yellow balls and there are 3 orange balls. We draw a ball from the bag, if the ball drawn is colored green or yellow we get \$0 payoff. If it is blue then we get a \$200 payoff. If the ball drawn is red, we get a \$300 payoff and \$500 if it is orange.
- (a) (5 points) What is our expected payoff of this experiment? **Hint:** Calculate expected value.
- (b) (5 points) What is the standard deviation and variance? **Hint:** $\sigma^2(X) = E[X^2] - (E[X])^2$

Answer: $x_i = 300, 200, 0, 0, 500$ for balls, Red, Blue, Green, Yellow and Orange respectively.
 $p_i = 9/39, 9/39, 9/39, 9/39, 3/39$ respectively.

a) Expected payoff = $E[X] = \sum x_i p_i = 300 * 9/39 + 200 * 9/39 + 0 * 9/39 + 0 * 9/39 + 500 * 3/39$
 $= 2000/13 = 153.84$ So, the expected payoff is \$153.84

b) $E[X^2] = \sum x_i^2 p_i = 90000 * 9/39 + 40000 * 9/39 + 250000 * 3/39 = 640000/13 = 49230.77$
 $E[X]^2 = 153.84^2 = 23666.74$
 $\sigma^2 = E[X^2] - (E[X])^2 = 49230.77 - 23666.74 = 25564.03 \implies \sigma = \sqrt{25564.03} = 159.89$
 $\sigma^2 = \text{Variance}$ $\sigma = \text{Standard Deviation}$

7. (10 points) Harry is 24 years old and plans to get his driver's license. The written knowledge test Harry has to give before the driving test is multiple choice where every question has y options. There are no negative marks for wrong answers and hence Harry chooses to guess an answer if he doesn't know the correct answer to the question. Let x be the probability that he knows the answer. The probability where Harry guesses at the answer and it will be correct is $\frac{1}{y}$. What is the probability that Harry knows the correct answer to a problem given that he has answered that problem correctly. (Assume that Harry does not leave any questions on the test unanswered. Also assume that for any question to which Harry knows the answer, his answers will always be correct.) **Hint:** Employ Bayes rule.

Answer: Let $p(k)$ be the probability of Harry knowing the answer = x . So, the probability of Harry guessing the answer, $p(g) = 1 - x$ because Harry only guesses if he doesn't know the answer.

Now, the probability of Harry guessing and getting the answer right, $p(g \cap c) = 1/y$
 So, $p(k \mid c) = \frac{p(k \cap c)}{p(c)} = \frac{p(c|k)*p(k)}{p(c)} = \frac{1*x}{p(c|k)*p(k)+p(c|g)*p(g)} = \frac{x}{1*x+i/y*(1-x)} = \frac{xy}{xy+1-x}$

8. (10 points) It is estimated that 40% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 98% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

Answer: Probability of an email being spam, $p(s) = 0.4$

So, probability of an email not being spam, $p(s') = 0.6$

Probability of the software giving a false positive, $p(+ve | s') = 0.05$

So, the probability of an email that was detected as spam, not being a spam is $p(s' | +ve) = \frac{p(+ve|s') \cdot p(s')}{p(+ve)}$
 $= \frac{0.05 \cdot 0.6}{p(+ve|s) \cdot p(s) + p(+ve|s') \cdot p(s')} = \frac{0.03}{0.98 \cdot 0.4 + 0.05 \cdot 0.6} = \frac{0.03}{0.422} = 0.071$

9. (10 points) A certain disease affects about 1 out of 10,000 people. There is a test to check whether the person has the disease. The test is quite accurate. In particular, we know that

- The probability that the test result is positive (suggesting the person has the disease), given that the person does not have the disease, is only 2 percent;
- The probability that the test result is negative (suggesting the person does not have the disease), given that the person has the disease, is only 1 percent.

A random person gets tested for the disease and the result is positive. What is the probability that the person has the disease?

Answer: The probability of a person getting the disease, $p(d) = 1/10000$

The probability of testing positive and not having the disease, $p(+ve|d') = 0.02$

The probability of testing negative and having the disease, $p(-ve|d) = 0.01$

So, the probability of testing positive given that the person has the disease, $p(+ve|d) = 1 - 0.01 = 0.99$

$p(d | +ve) = \frac{p(+ve|d) \cdot p(d)}{p(+ve)} = \frac{0.99 \cdot 1/10000}{p(+ve|d) \cdot p(d) + p(+ve|d') \cdot p(d')} = \frac{0.99}{10000 \cdot (0.99 \cdot 1/10000 + 0.02 \cdot (1 - 1/10000))} = \frac{0.99}{10000 \cdot (0.99 \cdot 1/10000 + 0.02 \cdot 0.9999)} = 0.0049$

Differential Calculus (25 points)

10. (6 points) **Limit Definition** of a function $F(x)$, is given as follows.

$$F(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

- (a) (3 points) Given a function $F(x) = \frac{f(x)}{g(x)}$ where $f(x)$ and $g(x)$ are also functions of x , the quotient rule of derivative of $F(x)$ (denoted $F'(x)$) is given as follows.

$$F'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$$

Use the limit definition of a function to derive the quotient rule of the derivative of function $F(x)$.

- (b) (3 points) Consider a function $f(x) = \sin(x)$. Show using the limit definition that $f'(x) = \cos(x)$.

Answer:

$$\begin{aligned} a) F(x) = \frac{f(x)}{g(x)} &\implies F'(x) = \lim_{h \rightarrow 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} = \lim_{h \rightarrow 0} \frac{\frac{f(x+h)g(x) - g(x+h)f(x)}{g(x+h)g(x)}}{h}, \text{ Adding and subtracting } f(x) \cdot g(x) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \cdot \frac{f(x+h)g(x) - g(x+h)f(x) + f(x)g(x) - f(x)g(x)}{g(x+h)g(x)} = \lim_{h \rightarrow 0} \frac{1}{h} \cdot \frac{f(x+h)g(x) - f(x)g(x)}{g(x+h)g(x)} - \frac{g(x+h)f(x) - f(x)g(x)}{g(x+h)g(x)} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \cdot \frac{f(x+h) - f(x)}{g(x+h)} - \frac{f(x)}{g(x+h)g(x)} \cdot (g(x+h) - g(x)) = \frac{1}{g(x)} \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} - \frac{f(x)}{g(x)^2} \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} \\ &= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}, \text{ hence proved.} \end{aligned}$$

$$\begin{aligned}
b) f(x) = \sin(x) &\implies f'(x) = \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin(x)}{h} = \lim_{h \rightarrow 0} \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h} \\
&= \lim_{h \rightarrow 0} \frac{-\sin(x)(1 - \cos(h))}{h} + \cos(x) \lim_{h \rightarrow 0} \frac{\sin(h)}{h}, \left(\lim_{h \rightarrow 0} \frac{\sin(h)}{h} = 1 \right) \implies f'(x) = \lim_{h \rightarrow 0} \frac{-\sin(x)(1 - \cos(h))}{h} + \cos(x) \\
(1 - \cos(x)) &= 2\sin^2(x/2) \implies f'(x) = \lim_{h \rightarrow 0} \frac{-\sin(x)(2\sin^2(h/2))}{h} = \lim_{h \rightarrow 0} \frac{\sin(h/2)}{h/2} \cdot \lim_{h \rightarrow 0} -\sin(x) \cdot \sin(h/2) + \cos(x) \\
&= -\sin(x)\sin(0) + \cos(x) = \cos(x), \text{ hence proved.}
\end{aligned}$$

11. (4 points) Calculate the equation of a tangent line given by a function $f(x) = x^4 + 5x^3 + 9x^2 - 3$ and report the result at $x = 1$ and $y = 12$. **Hint:** Calculate slope of function and employ point slope form of a line to get the tangent equation.

$$\text{Answer : } f(x) = x^4 + 5x^3 + 9x^2 - 3 \implies f'(x) = 4x^3 + 15x^2 + 18x \implies m = f'(1) = 4 + 15 + 18 = 37$$

$$y_1 = 12, x_1 = 1, m = 37. \text{ Using equation } y - y_1 = m(x - x_1) \implies y - 12 = 37(x - 1) \implies y = 37x - 25$$

12. (4 points) Consider a function $f(x) = e^{2x}$ and a function $g(x) = \frac{x^2}{x-1}$. Let $F(x) = f(g(x))$, find an expression for $F'(x)$. **Hint:** Employ chain rule. Answer does not need to be simplified after FULL application of derivatives (the point is to test your application of derivatives not your algebra). Please enumerate the derivative rules you have employed.

$$\text{Answer : Let } g(x) = u, F(x) = f(u) = e^{2u} \implies \frac{d}{du}f(u) = 2e^{2u}, u = \frac{x^2}{x-1} \implies \frac{du}{dx} = \frac{2x(x-1) - 1(x^2)}{(x-1)^2} \text{ Quotient rule}$$

$$= \frac{x^2 - 2x}{(x-1)^2} \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \text{ chain rule} \implies \frac{dy}{dx} = 2e^{2u} \cdot \frac{x^2 - 2x}{(x-1)^2} = 2e^{\frac{x^2}{x-1}} \frac{x^2 - 2x}{(x-1)^2}$$

13. (6 points) Given $F(x)$ in the following problems, find the derivative $F'(x)$. **Note:** Answer does not need to be simplified after FULL application of derivatives (the point is to test your application of derivatives not your algebra). Please enumerate the derivative rules you have employed per question.

(a) (2 points) $F(x) = \frac{\sqrt{x+2x}}{7x-4x^2}$

(b) (2 points) $F(x) = (1 + \sqrt{x^3})(\frac{1}{x^3} - 2\sqrt[3]{x})$

(c) (2 points) $F(x) = (2x^2 + 1)^3(3x^3 - 2)^2$

$$\text{Answer : a) } F(x) = \frac{\sqrt{x+2x}}{7x-4x^2} \quad F'(x) = \frac{(\frac{1}{2\sqrt{x}} + 2)(7x - 4x^2) - (7 - 8x)(\sqrt{x+2x})}{(7x - 4x^2)^2} \text{ Quotient rule}$$

$$b) F(x) = (1 + \sqrt{x^3})\left(\frac{1}{x^3} - 2\sqrt[3]{x}\right) F'(x) = \frac{3\sqrt{x}}{2}\left(\frac{1}{x^3} - 2\sqrt[3]{x}\right) + \left(\frac{-3}{x^4} - \frac{2}{3\sqrt[3]{x^2}}\right)(1 + \sqrt{x^3}) \text{ Product rule}$$

$$c) F(x) = (2x^2 + 1)^3(3x^3 - 2)^2 F'(x) = 12x(2x^2 + 1)^2(3x^3 - 2)^2 + 12x(3x^3 - 2)(2x^2 + 1)^3 \text{ Chain and Product rules}$$

14. (2 points) Calculate the partial derivatives with respect to each variable of the function $F(\cdot)$ in each of the following cases.

$$(a) \text{ (1 point) } F(x, y) = \frac{x^3 y^2 - 2x^2 + 5y}{e^x}$$

$$(b) \text{ (1 point) } F(x, y, z) = y^2 \ln(x + 2y) - \ln(3z)(x^3 + y^2 - 4z)$$

$$\text{Answer : } a) \frac{\delta}{\delta x} F(x, y) = \frac{e^x(3x^2 y^2 - 4x) - e^x(x^3 y^2 - 2x^2 + 5y)}{e^{2x}}, \quad \frac{\delta}{\delta y} F(x, y) = \frac{2x^3 y + 5}{e^x}$$

$$b) \frac{\delta}{\delta x} F = \frac{y^2}{x + 2y} - \ln(3z)(2x^2), \quad \frac{\delta}{\delta y} F = 2y \ln(x + 2y) + \frac{2y^2}{x + 2y} - \ln(3z)(2y), \quad \frac{\delta}{\delta z} F = \frac{-x^3}{z} - \frac{y^2}{z} + 4\ln(3z) + 4$$

15. (3 points) The table below depicts rainfall depth (mm) in New York City beginning 7 PM on Sept. 19th 2008. **Hint:** In each of the following questions, employ numerical differentiation techniques.

Hours after 7 PM	0	5	10	15	20	25	30
Rainfall Depth (mm)	100	250	300	370	325	400	450

Table 2: Rainfall Depth in New York City Sept 19th 2008, at and after 7 PM.

- (a) (1 point) Use central difference to approximate the slope of rainfall depth using numerical differentiation at hour 12.5 and at hour 22.5.
- (b) (1 point) Calculate the slope of rainfall depth at hour 15 using forward numerical differentiation.
- (c) (1 point) Calculate the slope of rainfall depth at hour 15 using backward numerical differentiation.

$$\text{Answer : } a) h = 2.5, f'(12.5) = \frac{f(12.5 + 2.5) - f(12.5 - 2.5)}{2 * 2.5} = \frac{370 - 300}{5} = \frac{70}{5} = 14$$

$$f'(22.5) = \frac{f(22.5 + 2.5) - f(22.5 - 2.5)}{5} = \frac{f(25) - f(20)}{5} = \frac{400 - 325}{5} = \frac{75}{5} = 15$$

$$b) h = 5, f'(15) = \frac{f(15 + 5) - f(15)}{5} = \frac{f(20) - f(15)}{5} = \frac{325 - 370}{5} = \frac{-45}{5} = -9$$

$$c) h = 5, f'(15) = \frac{f(15) - f(15 - 5)}{5} = \frac{f(15) - f(10)}{5} = \frac{370 - 300}{5} = \frac{70}{5} = 14$$