

Instructor: Nikhil Muralidhar
April 6, 2025

CS 556-C: 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)

1. (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.

Finding the eigenvalues

$$\det \begin{bmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{bmatrix} = (1-\lambda)(3-\lambda) - 8 = \lambda^2 - 4\lambda - 5 = 0$$

$$\lambda = \frac{4 \pm \sqrt{16+20}}{2} = \frac{4 \pm 6}{2}$$

$$\lambda_1 = 5, \lambda_2 = -1$$

Finding the eigenvectors

For $\lambda_1 = 5$:

$$\begin{bmatrix} 1-5 & 2 \\ 4 & 3-5 \end{bmatrix} = \begin{bmatrix} -4 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 = -4v_1 + 2v_2 = 0$$

$$v_2 = 2v_1$$

$$v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

For $\lambda_2 = -1$:

$$\begin{bmatrix} 1+1 & 2 \\ 4 & 3+1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 = 2v_1 + 2v_2 = 0$$

$$v_1 = -v_2$$

$$v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

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

$$P = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$$

2. (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.

(a) Finding the eigenvalues of $A^T A$:

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

$$\det \begin{bmatrix} 13-\lambda & 12 \\ 12 & 13-\lambda \end{bmatrix} = (13-\lambda)^2 - 144 = (\lambda-1)(\lambda-25) = 0$$

$$\lambda_1 = 25, \lambda_2 = 1$$

$$\sigma_1 = \sqrt{25} = 5 \quad \sigma_2 = \sqrt{1} = 1$$

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

(b) Finding the eigenvectors of $A^T A$:

For $\lambda_1 = 25$:

$$\begin{bmatrix} -12 & 12 \\ 12 & -12 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$v_1 + v_2 = 0$$

$$v_1 = -v_2$$

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 1$:

$$\begin{bmatrix} 12 & 12 \\ 12 & 12 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$v_1 + v_2 = 0$$

$$v_1 = v_2$$

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$AA^T = A^T A \text{ so } U = V$$

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

(c) Constructing the SVD: $A = U\Sigma V^T$

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$

$V^T = V$ so

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

The final SVD is:

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

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).

- (a) (1 point) What is U1's 'estimated' rating of the movie *Amelie*?
 - (b) (2 point) What is the *strength* of concept 1 in the SVD decomposition?
 - (c) (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)?
 - (d) (1 point) What is the movie with the overall highest rating (i.e., if actual ratings exist, use them if not use estimated ratings)?
 - (e) (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).
4. (7 points) Calculate the principle component analysis of the Iris dataset using the built-in method from Scikit-learn and showcase the following.
- (a) (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.
 - (b) (1 point) What is the total percentage of variance captured by the first 2 components of PCA?
 - (c) (1 point) What is the *strength* of each of the two principal components?
 - (d) (1 point) What is the magnitude of each of the two principal components?
 - (e) (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)

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.
 - $\mu = 180g$
 - $\sigma = 60g$
 - Jane does not like apples with weight $\leq 100g$

The Z-score for 100g is:

$$Z = \frac{X - \mu}{\sigma} = \frac{100 - 180}{60} = -1.33$$

$$P(Z \leq -1.33) = 0.09176 \text{ or } 9.18\%$$
 (a) 9.18% of apples are not liked by Jane.
 - (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. Out of all the apples liked by Jane, what is the probability that the weight of a randomly selected apple 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 $\geq q_3$) - q_3 is the 75th percentile of the distribution.
 - Liked apples are $> 100g$.
 - What is the probability that Jane likes an apple AND that the weight of the apple is $\geq q_3$?

$$P(X > 100) = 1 - P(X \leq 100) = 1 - 0.0918 = 0.9082 \text{ from part a}$$

$$P(X \leq q_3) = P(X \leq 100) + 0.75 \times P(X > 100) = 0.0918 + 0.75 \times 0.9082 = 0.7729$$

Need to find q_3 such that $P(X \leq q_3 | X > 100) = 0.75$

$$P(X \leq q_3) = P(X \leq 100) + 0.75 \times P(X > 100) = 0.0918 + 0.75 \times 0.9082 = 0.7729$$

From the Z-table, $P(Z \leq 0.75)$ is approximately 0.7734.

$$Z = \frac{q_3 - 180}{60} = 0.75$$

$$q_3 = \mu + Z\sigma = 180 + 0.75 \times 60 = 180 + 45 = 225g$$

By definition q_3 represents the upper quartile, i.e., top 25%, of weight of apples liked by Jane.

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 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. There are 36 balls in total

There are 9 red, blue green and yellow balls and thus a probability of $\frac{9}{39}$

There are 3 orange balls each with a probability of $\frac{3}{39}$

$$E[X] = (\$0 \times P(\text{Green})) + (\$0 \times P(\text{Yellow})) + (\$200 \times P(\text{Blue})) + (\$300 \times P(\text{Red})) + (\$500 \times P(\text{Orange}))$$

$$E[X] = (\$0 \times \frac{9}{39}) + (\$0 \times \frac{9}{39}) + (\$200 \times \frac{9}{39}) + (\$300 \times \frac{9}{39}) + (\$500 \times \frac{3}{39})$$

$$E[X] = 0 + 0 + \$46.15 + \$69.23 + \$38.46$$

$$E[X] = \$153.84$$

- (b) (5 points) What is the standard deviation and variance? **Hint:** $\sigma^2(X) = E[X^2] - (E[X])^2$

Calculate $E[X^2]$

$$E[X^2] = (\$0^2 \times P(\text{Green})) + (\$0^2 \times P(\text{Yellow})) + (\$200^2 \times P(\text{Blue})) + (\$300^2 \times P(\text{Red})) + (\$500^2 \times P(\text{Orange}))$$

$$E[X^2] = (\$0^2 \times \frac{9}{39}) + (\$0^2 \times \frac{9}{39}) + (\$200^2 \times \frac{9}{39}) + (\$300^2 \times \frac{9}{39}) + (\$500^2 \times \frac{3}{39})$$

$$E[X^2] = 0 + 0 + \frac{360000}{39} + \frac{810000}{39} + \frac{750000}{39}$$

$$E[X^2] = \frac{360000 + 810000 + 750000}{39}$$

$$E[X^2] = \frac{1920000}{39}$$

$$E[X^2] = \$49230.77$$

Calculate variance

$$\sigma^2(X) = E[X^2] - (E[X])^2 = \frac{1920000}{39} - (\frac{6000}{39})^2$$

$$(\frac{6000}{39})^2 = \frac{36000000}{1521}$$

$$\frac{1920000}{39} = \frac{1920000 \times 39}{39 \times 39} = \frac{74880000}{1521}$$

$$\sigma^2(X) = \frac{74880000 - 36000000}{1521} = \frac{38880000}{1521}$$

$$\text{Variance } \sigma^2(X) = 25562.13$$

$$\text{Standard Deviation } \sigma(X) = \sqrt{25562.13} = 159.5$$

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.

K Harry knows the right answer

C Harry answers the question correctly

$$P(K) = x$$

$$P(\text{not}K) = 1 - x$$

$$\begin{aligned}
P(C|notK) &= 1/y \\
P(C|K) &= 1 \\
P(K|C) &= P(K)P(C|K)/P(C) \\
P(C) &= P(K)P(C|K) + P(notK)P(C|notK) \\
P(C) &= (1 * x) + (1/y)(1 - x) \\
P(C) &= x + (1/y)(1 - x) \\
P(K|C) &= \frac{1 * x}{x + (1/y)(1 - x)} \\
P(K|C) &= \frac{xy}{xy + 1 - x}
\end{aligned}$$

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?

Using Bayes' Theorem.

$$P(\text{Not Spam} \text{---} \text{Detected as Spam}) = P(\text{Detected as Spam} \text{---} \text{Not Spam})P(\text{Not Spam})/P(\text{Detected as Spam})$$

$$P(\text{Detected as Spam} \text{---} \text{Not Spam}) = 0.05$$

$$P(\text{Not Spam}) = 0.6$$

$$P(\text{Detected as Spam}) = P(\text{Detected as Spam} \text{---} \text{Spam})P(\text{Spam}) + P(\text{Detected as Spam} \text{---} \text{Not Spam})P(\text{Not Spam})$$

$$P(\text{Detected as Spam}) = 0.98 * 0.4 + 0.05 * 0.6$$

$$P(\text{Detected as Spam}) = 0.392 + 0.03$$

$$P(\text{Detected as Spam}) = 0.422$$

$$P(\text{Not Spam} \text{---} \text{Detected as Spam}) = 0.05 * 0.6 / 0.422$$

$$P(\text{Not Spam} \text{---} \text{Detected as Spam}) = 0.071 \text{ or } 7.11\%$$

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?

$$P(D) = 1/10000 = 0.0001$$

$$P(\text{not } D) = 1 - P(D) = 0.9999$$

$$\text{False Positive} = P(\text{Positive} \text{---} \text{not } D) = 0.02$$

$$\text{False Negative} = P(\text{Negative} \text{---} D) = 0.01$$

$$P(\text{Positive} \text{---} D) = 1 - P(\text{Negative} \text{---} D) = 0.99$$

$$P(\text{Positive}) = P(\text{Positive} \text{---} D)P(D) + P(\text{Positive} \text{---} \text{not } D)P(\text{not } D)$$

$$P(\text{Positive}) = 0.99 * 0.0001 + 0.02 * 0.9999 = 0.000099 + 0.019998 = 0.020097$$

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 $f'(x) = \cos(x)$.

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.

$$f'(1) = 4(1)^3 + 15(1)^2 + 18(1) = 4 + 15 + 18 = 37 \text{ Slope of tangent line} = 37$$

$$\text{Point slope form of a line: } y - y_1 = m(x - x_1)$$

where m is the slope of the line and (x_1, y_1) is a point on the line.

Substituting the values we have: $y - 12 = 37(x - 1)$

$$y - 12 = 37x - 37$$

$$y = 37x - 25$$

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.

Applying the chain rule: $F'(x) = f'(g(x)) \times g'(x)$

$$f'(x) = \frac{d}{dx} e^{2x} = e^{2x} \times \frac{d}{dx} 2x = e^{2x} \times 2$$

$$g'(x) = \frac{d}{dx} \frac{x^2}{x-1}$$

Applying the quotient rule where $u = x^2$ and $v = x - 1$

$$u' = 2x$$

$$v' = 1$$

$$g'(x) = \frac{(x-1)(2x) - (x^2)(1)}{(x-1)^2}$$

$$g'(x) = \frac{2x^2 - 2x - x^2}{(x-1)^2}$$

$$g'(x) = \frac{x^2 - 2x}{(x-1)^2}$$

Substituting the values of $f'(g(x))$ and $g'(x)$ into the chain rule gives us:

$$F'(x) = f'(g(x)) \times g'(x)$$

$$F'(x) = e^{2g(x)} \times 2 \times \frac{(x^2 - 2x)}{(x-1)^2}$$

$$F'(x) = 2e^{2\frac{x^2}{x-1}} \times \frac{(x^2 - 2x)}{(x-1)^2}$$

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

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

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

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

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

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.