

CSE4088 Introduction to Machine Learning

Homework 1 - Prerequisite Concepts

The goal of this homework is to refresh your previous knowledge on mathematical and other background concepts. If you have difficulties in answering the below questions, you are suggested to refer to the resources listed at the end of the homework to fill in this background.

I. Basic Questions

1. Vectors and Matrices (10 points)

Consider the matrix X and the vectors \mathbf{y} and \mathbf{z} below:

$$X = \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 9 \\ 8 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$$

1. What is the inner product of the vectors \mathbf{y} and \mathbf{z} ? (this is also sometimes called the *dot product*, and is sometimes written as $\mathbf{y}^T \mathbf{z}$)
2. What is the product $X\mathbf{y}$?
3. Is X invertible? If so, give the inverse, and if no, explain why not.
4. What is the rank of X ?

2. Calculus (10 points)

1. If $y = 4x^3 - x^2 + 7$ then what is the derivative of y with respect to x ?
2. If $y = \tan(z)x^{6z} - \ln(\frac{7x+z}{x^4})$, what is the partial derivative of y with respect to x ?

3. Probability and Statistics (20 points)

Consider a sample of data $S = \{0, 1, 1, 0, 0, 1, 1\}$ created by flipping a coin x seven times, where 0 denotes that the coin turned up heads and 1 denotes that it turned up tails.

1. What is the sample mean for this data?
2. What is the sample variance for this data?
3. What is the probability of observing this data, assuming it was generated by flipping a biased coin with $p(x = 1) = 0.7, p(x = 0) = 0.3$.
4. Note that the probability of this data sample would be greater if the value of $p(x = 1)$ was not 0.7, but instead some other value. What is the value that maximizes the probability of the sample S ? Please justify your answer.

5. Consider the following joint probability table where both A and B are binary random variables:

A	B	$P(A, B)$
0	0	0.1
0	1	0.4
1	0	0.2
1	1	0.3

- (a) What is $P(A = 0, B = 0)$?
- (b) What is $P(A = 1)$?
- (c) What is $P(A = 0|B = 1)$?
- (d) What is $P(A = 0 \vee B = 0)$?

4. Big-O Notation (20 points)

For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, both, or neither. Briefly justify your answers.

1. $f(n) = \frac{n}{2}$, $g(n) = \log_2(n)$.
2. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
3. $f(n) = n^{100}$, $g(n) = 100^n$.

II. Medium-Level Questions

5. Algorithm (5 points)

Divide and Conquer: Assume that you are given a sorted array with n integers in the range $[-10, +10]$. Note that some integer values may appear multiple times in the array. Additionally, you are told that somewhere in the array the integer 0 appears exactly once. Provide an algorithm to locate the 0 which runs in $O(\log(n))$. Explain your algorithm in words, describe why the algorithm is correct, and justify its running time.

6. Probability and Random Variables (15 points)

6.1 Probability

State true or false. Here Ω denotes the sample space and A^c denotes the complement of the event A .

1. For any $A, B \subseteq \Omega$, $P(A|B)P(B) = P(B|A)P(A)$.
2. For any $A, B \subseteq \Omega$, $P(A \cup B) = P(A) + P(B) - P(A|B)$.
3. For any $A, B, C \subseteq \Omega$ such that $P(B \cup C) > 0$, $\frac{P(A \cup B \cup C)}{P(B \cup C)} \geq P(A|B \cup C)P(B \cup C)$.
4. For any $A, B \subseteq \Omega$ such that $P(B) > 0$, $P(A^c) > 0$, $P(B|A^c) + P(B|A) = 1$.
5. For any n events $\{A_i\}_{i=1}^n$, if $P(\bigcap_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, then $\{A_i\}_{i=1}^n$ are mutually independent.

6.2 Discrete and Continuous Distributions

Match the distribution name to its probability density / mass function. Below, $|x| = k$.

- | | |
|-----------------|--|
| | (f) $f(x; \Sigma, \mu) = \frac{1}{\sqrt{(2\pi)^k \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$ |
| | (g) $f(x; n, \alpha) = \binom{n}{x} \alpha^x (1 - \alpha)^{n-x}$ for $x \in \{0, \dots, n\}$; 0 otherwise |
| (a) Laplace | (h) $f(x; b, \mu) = \frac{1}{2b} \exp\left(-\frac{ x-\mu }{b}\right)$ |
| (b) Multinomial | (i) $f(x; n, \alpha) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k \alpha_i^{x_i}$ for $x_i \in \{0, \dots, n\}$ and $\sum_{i=1}^k x_i = n$; 0 otherwise |
| (c) Poisson | (j) $f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ for $x \in (0, +\infty)$; 0 otherwise |
| (d) Dirichlet | (k) $f(x; \alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i-1}$ for $x_i \in (0, 1)$ and $\sum_{i=1}^k x_i = 1$; 0 otherwise |
| (e) Gamma | (l) $f(x; \lambda) = \lambda^x \frac{e^{-\lambda}}{x!}$ for all $x \in \mathbb{Z}^+$; 0 otherwise |

6.3 Mean and Variance

- Consider a random variable which follows a Binomial distribution: $X \sim \text{Binomial}(n, p)$.
 - What is the mean of the random variable?
 - What is the variance of the random variable?
- Let X be a random variable and $\mathbb{E}[X] = 1$, $\text{Var}(X) = 1$. Compute the following values:
 - $\mathbb{E}[3X]$
 - $\text{Var}(3X)$
 - $\text{Var}(X + 3)$

6.4 Mutual and Conditional Independence

- If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- If X and Y are independent random variables, show that $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
Hint: $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$
- If we roll two dice that behave independently of each other, will the result of the first die tell us something about the result of the second die? If, however, the first die's result is a 1, and someone tells you about a third event — that the sum of the two results is even — then given this information is the result of the second die independent of the first die?

6.5 Law of Large Numbers and the Central Limit Theorem

Provide one line justifications.

- Suppose we simultaneously flip two independent fair coins (i.e., the probability of heads is $1/2$ for each coin) and record the result. After 40,000 repetitions, the number of times the result was two heads is close to 10,000. (Hint: calculate how close.)
- Let $X_i \sim \mathcal{N}(0, 1)$ and $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, then the distribution of \bar{X} satisfies

$$\sqrt{n}\bar{X} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

7. Linear Algebra (10 points)

7.1 Norms

Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $\|\mathbf{x}\|_1 \leq 1$ (Recall that $\|\mathbf{x}\|_1 = \sum_i |x_i|$)
2. $\|\mathbf{x}\|_2 \leq 1$ (Recall that $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$)
3. $\|\mathbf{x}\|_\infty \leq 1$ (Recall that $\|\mathbf{x}\|_\infty = \max_i |x_i|$)

7.2 Geometry

Prove that these are true or false. Provide all steps.

1. The smallest Euclidean distance from the origin to some point \mathbf{x} in the hyperplane $\mathbf{w}^T \mathbf{x} + b = 0$ is $\frac{|b|}{\|\mathbf{w}\|_2}$.
2. The Euclidean distance between two parallel hyperplane $\mathbf{w}^T \mathbf{x} + b_1 = 0$ and $\mathbf{w}^T \mathbf{x} + b_2 = 0$ is $\frac{|b_1 - b_2|}{\|\mathbf{w}\|_2}$ (Hint: you can use the result from the last question to help you prove this one).

8. Programming Skills (10 points)

Sampling from a distribution. For each question, submit a scatter plot (you will have 5 plots in total). Make sure the axes for all plots have the same limits. (Hint: You can save a Matlab figure as a pdf, and then use includegraphics to include the pdf in your latex file.)

1. Draw 100 samples $\mathbf{x} = [x_1, x_2]^T$ from a 2-dimensional Gaussian distribution with mean $(0, 0)^T$ and identity covariance matrix, i.e., $p(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right)$, and make a scatter plot (x_1 vs. x_2). For each question below, make each change separately to this distribution.
2. Make a scatter plot with a changed mean of $(1, -1)^T$.
3. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.
4. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & 0.2 \\ 0.2 & 2 \end{pmatrix}$.
5. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & -0.2 \\ -0.2 & 2 \end{pmatrix}$.

Submission Instructions:

1. Each student should submit his/her own homework. You can discuss the questions with your friends, but you must write your own solutions and code. Group work is not allowed. You can not exchange any written material, code or pseudocode. This also includes material found on the web.
2. Write a detailed report using Word/Latex and convert it to a pdf file. The report must include explanations about each part in each question. You should solve some of the questions by pencil and paper. Then, you can scan your solution and insert it to your report (you can use free Mobile applications such as Adobe Scan).
3. Explain how your scripts and functions work, i.e., which parts of your functions/scripts accomplish which task and how it is accomplished. Include the outputs of your functions and figures to your report. Each figure should have a caption and should be explained in the text.

4. You can use MATLAB or Python for programming assignments. Don't forget to put detailed comments into your functions/scripts to explain what your code is doing. Also indicate the inputs and outputs in the comment section.
5. Combine your report and codes into a single zip file. Plots and figures should go into the report.
6. Name your zip file as "CSE4088_name_surname_hw_no.zip". For example, a student whose name is Ayşe Çalışkan will name her file as: "CSE4088_ayse_caliskan_hw1.zip" for the first homework. Also, write your name, surname and student number as comments at the beginning of your codes.
7. Submit your homework via the class web page at classroom.google.com before the deadline.
8. Late submissions will lose 10 points for each day after the deadline.

Resources:

Probability and Statistics:

- Lecture Notes: http://www.cs.cmu.edu/~aarti/Class/10701/recitation/prob_review.pdf
- Book: Probability and Statistics for Engineers and Scientists, Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye
https://fac.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineers_and_scientists.pdf
- MIT OpenCourseware – Intro. to Probability and Statistics
<https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/>

Linear Algebra:

- Short video lectures by Prof. Zico Kolter: <http://www.cs.cmu.edu/~zkolter/course/linalg/outline.html>
- Handout associated with above video: http://www.cs.cmu.edu/~zkolter/course/linalg/linalg_notes.pdf
- Book: Gilbert Strang. Linear Algebra and its Applications. HBJ Publishers.
- MIT OpenCourseware – Linear Algebra <https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/>

Calculus:

- MIT OpenCourseware – Single Variable Calculus: <https://ocw.mit.edu/courses/mathematics/18-01-single-variable-calculus-fall-2006/>
- MIT OpenCourseware – Multivariable Calculus: <https://ocw.mit.edu/courses/mathematics/18-02sc-multivariable-calculus-fall-2010/>
- MIT OpenCourseware – Highlights of Calculus: <https://ocw.mit.edu/resources/res-18-005-highlights-of-calculus-spring-2010/index.htm>

MATLAB Tutorial:

- https://www.mathworks.com/help/pdf_doc/matlab/matlab_prog.pdf
- http://www.math.mtu.edu/_msgocken/intro/intro.pdf.
- <http://ubcmatlabguide.github.io/>.

Big-O notation:

- http://www.stat.cmu.edu/_cshalizi/uADA/13/lectures/app-b.pdf
- http://www.cs.cmu.edu/_avrim/451f13/recitation/rec0828.pdf