CIS 4190/5190: Applied Machine Learning	Fall 2025

Homework 3

Handed Out: October 22 — *Due: 7:59 pm November 5*

- You are encouraged to format your solutions using LaTeX. Handwritten solutions are permitted, but remember that you bear the risk that we may not be able to read your work and grade it properly — we will not accept post hoc explanations for illegible work. You will submit your solution manuscript for written HW 3 as a single PDF file.
- Familiarize yourself with the "Human and AI Assistance in Homework" Policy included in the administrivia slides from the first lecture. `https://www.seas.upenn.edu/~cis5190/fall2025/schedule.html`
- The homework is **due at 8 PM** on the due date. We will be using Gradescope for collecting the homework assignments. Please submit your solution manuscript as a PDF file via Gradescope. Post on Ed Discussion and contact the TAs if you are having technical difficulties in submitting the assignment.
- Make sure to assign pages to each question when submitting homework to Gradescope. The TA may deduct 0.2 points per sub-question if a page is not assigned to a question.
- Items marked [**5190 Only**] are mandatory for students enrolled in CIS 5190, and optional for CIS 4190. More information on the administrivia slides.

1 Written Questions

Note: You do not need to show work for multiple choice questions. If formatting your answer in LaTeX, use our LaTeX template `hw_template.tex` (This is a read-only link. You'll need to make a copy before you can edit. Make sure you make only private copies.).

1. [k-Means] (10 pts) Consider the k-means clustering algorithm:

   (a) [8 pts] Work through the K-Means clustering algorithm for a dataset with 4 samples, with $K = 2$, and using the L_2 distance. The samples in the dataset are: $A = (2, 3), B = (4, 6), C = (5, 1)$, and $D = (10, 12)$. The initial centroids are chosen as: $(6, 9)$ for cluster 1 and $(8, 4)$ for cluster 2. Recall that in each iteration of K-Means, two things happen: first, cluster assignments are updated, and second, cluster centroids are updated. Work through two such iterations. Report results for each iteration as:

   - A: $d(A, 1)$, $d(A, 2)$
   - B: $d(B, 1)$, $d(B, 2)$
   - C: $d(C, 1)$, $d(C, 2)$
   - D: $d(D, 1)$, $d(D, 2)$
   - cluster 1 members: A, B, etc.

- cluster 1 updated centroid: (x, y)
- cluster 2 members: A, B, etc.
- cluster 2 updated centroid: (x, y)

where $d(S, c)$ is the L_2 distance from sample S to the cluster c centroid. You should write all your answers as a single number or radical. It is fine to leave decimals in the radical.

(b) [2 pts] Consider what happens when $K = 4$. Assume we run K-Means with an initialization that achieves the global minimum loss. What will be the coordinates of the four learned centroids? (hint: you don't need to calculate anything like in part a)

2. [PCA] (10 pts) *Note: You are expected to work out the entire question by hand, and not use any libraries/packages. Plots can be drawn manually/through drawing tools, but the labels and lines in the plots have to figured out manually.*

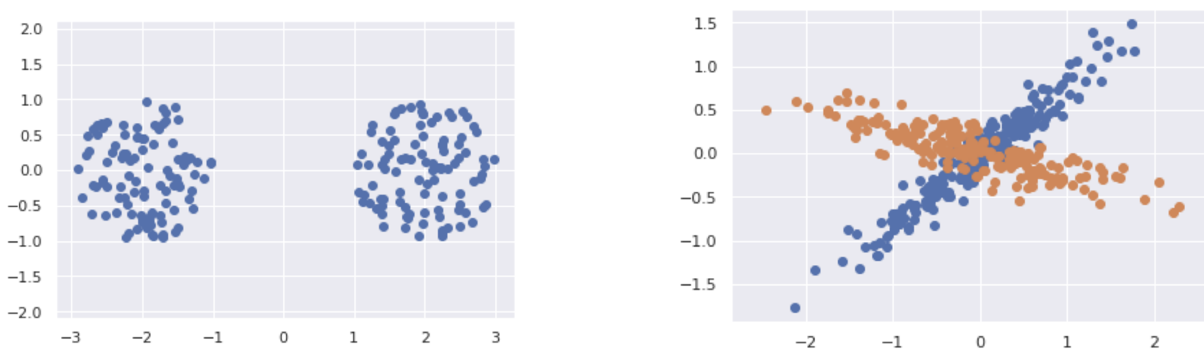


Figure 1: Draw the first and second principal components on each plot.

(a) [5 pts] Principal component analysis is a dimensionality reduction method that projects a dataset into its most variable components. You are given the following 2D datasets, draw the first and second principal components on each plot in Fig 1. *Note: Different colors are purely for visualization.*

(b) [5 pts] **[5190 Only]** Assume we are given a dataset for which the eigenvalues of the covariance matrix are: (2.1, 1.8, 1.3, 0.9, 0.4, 0.2, 0.15, 0.02, 0.001). What is the smallest value of K (dimension after reduction) we can use if we want to retain 75% of the variance (sum of all the variances in value) using the first principal components? Justify your answer; you may use a 4-function calculator for computations.

3. [PCA] (13 pts) *Note: You are expected to work out the entire question by hand, and not use any libraries/packages. Plots can be drawn manually/through drawing tools, but the labels and lines in the plots have to figured out manually.*

Bob wants to transmit the following set of four two-dimensional coordinates to his friend.

$$X = \begin{bmatrix} 4 & 1 \\ 2 & 3 \\ 5 & 4 \\ 1 & 0 \end{bmatrix}$$

However, due to transmission bandwidth, he is contrained to send only four one-dimensional coordinates. He learns about the PCA algorithm and wants to apply it.

(a) [8 pts] Find the unit-vector principal components of X. Given that Bob is constrained to send four 1-D coordinates, which principal component would you suggest him to pick, and why? Show your work.

(b) [5 pts] Bob plots the coordinates of X as in Figure 1. To obtain the 1-D transformation of his 2-D coordinates, he sketches the direction of the principal component and projects the four 2-D coordinates on this principle component. Show how this plot would look like. Label each of the projected points along with the value of the principal coordinate (note that these labels should be 1-D points).

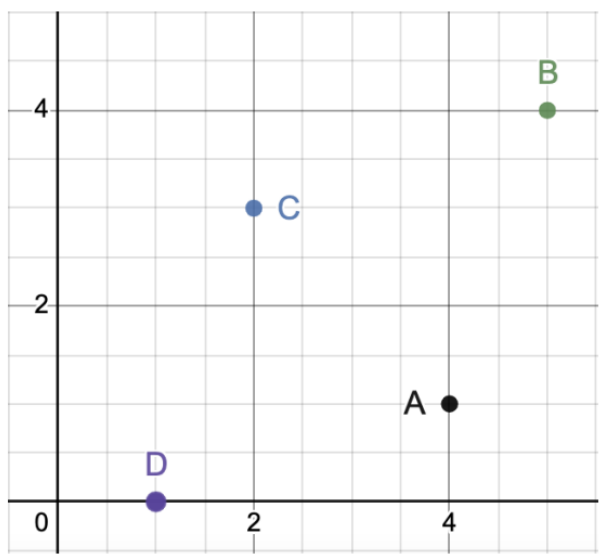


Figure 2: Bob's plot

4. [Image Filtering/Convolution] (9 pts) We discussed the use of convolution filters for images briefly in class, mostly in the context of CNN. However, before CNNs became popular, convolution filters had already been an essential part of signal processing and computational photography. You will probably be surprised by how many features in Photoshop or Lightroom can be easily implemented with the correct choice of convolution filter(s). In this question, we will take a look at a few common types of convolution

filters for images, and visualize how they would transform the original image. Let's take the following image for example. This is a gray-scale image, where each pixel can be represented by a value between $[0, 1]$, where 0 is black and 1 is white. The gray-scale image itself can be represented by a matrix of shape $height \times width$, and we are going to apply 3×3 convolution filters to the matrix. Assume the bias parameter is set to 0 for all these convolution filters.

For the following sub-questions, you will be given a convolution filter, and a few transformed image. Your task is to pick the one that corresponds to the given filter.

(a) [3 pts] Consider the following filter:

$$X = \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

If we apply X as a convolution filter to the original image, which of the following transformed image will we see? **Briefly justify your choice using one or two sentences.**

- A. 
- B. 
- C. 
- D. 

(b) [3 pts] Now consider this following filter:

$$X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

If we apply X as a convolution filter to the original image, which of the following transformed image will we see? **Briefly justify your choice using one or two sentences.**

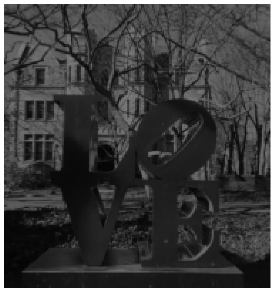
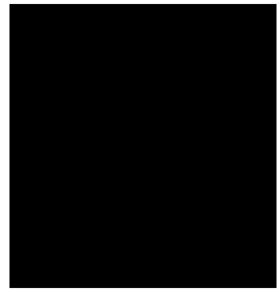
- A. 
- B. 
- C. 
- D. 

(c) [3 pts] Now let's look at a more challenging example:

$$X = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

If we apply X as a convolution filter to the original image, which of the following transformed image will we see? **Briefly justify your choice using one or two sentences.**

- A. 
- C. 

- B. 
- D. 

5. [CNNs] (8 pts) Consider the following 3-layer neural network.

    i. **Input**: RGB image $320 \times 320 \times 3$

    ii. **Conv2d**: 5×5 kernel, Stride=1, Padding=2, Out channels=16

    iii. **ReLU**

    iv. **MaxPool2d**: kernel: 2x2, Stride 2, Padding 0

    v. **Conv2d**: Out channels: 8, kernel: 3x3, Stride 1, Padding 1

    vi. **ReLU**

    vii. **MaxPool2d**: kernel: 2x2, Stride 2, Padding 0

    viii. **Conv2d**: Out channels: 4, kernel: 3x3, Stride 1, Padding 1

    ix. **ReLU**

    x. **MaxPool2d**: kernel: 2x2, Stride 2, Padding 0

    If the input image is 320×320 pixels with RGB channels (3 channels), compute the following:

    (a) [2 pts] Output size dimension 1

    (b) [2 pts] Output size dimension 2

    (c) [2 pts] Output size dimension 3

    (d) [2 pts] Number of learnable parameters