CIS 4190/5190: Applied Machine Learning — Fall 2024

Homework 3

Handed Out: October 23 — *Due: 7:59 pm November 6*

- 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.
- The homework is **due at 7:59 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.

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 `hw3_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. [PCA] (8 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.*

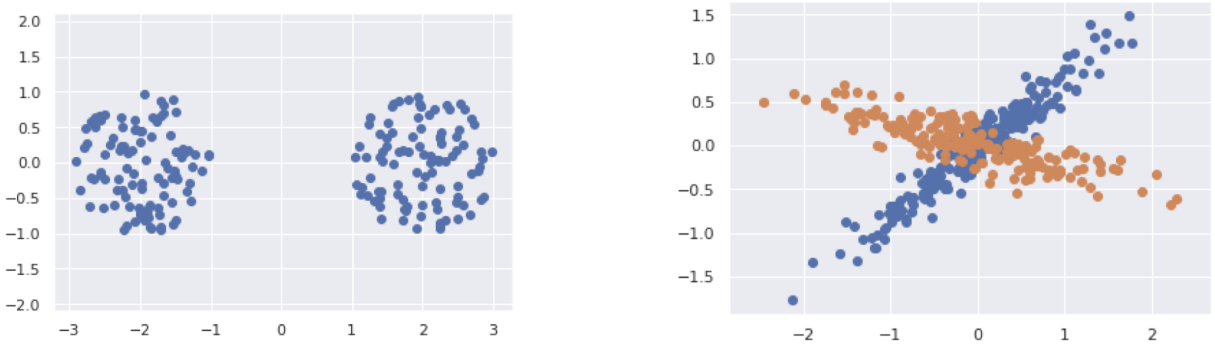


Figure 1: Draw the first and second principal components on each plot.

(a) [4 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.

(b) [2 pts] 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.

(c) [2 pts] Assume we apply PCA to a matrix $X \in \mathbb{R}^{N \times M}$ and obtain a set of PCA features. We divide this set into two, $Z1$ and $Z2$. The first set, $Z1$, corresponds to the top principal components. The second set, $Z2$, corresponds to the remaining principal components.
Which one from the following options is more common in the training data?

a) a point with large feature values in $Z1$ and small feature values in $Z2$
b) a point with small feature values in $Z1$ and small feature values in $Z2$
c) a point with large feature values in $Z1$ and large feature values in $Z2$
d) a point with small feature values in $Z1$ and large feature values in $Z2$

2. (8 pts) [k-Means] 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.

3. [Image Filtering/Convolution] (6 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 $width * height$, 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. Your TAs have created a skeleton colab notebook, where you can implement + test out these filters with different images. https://colab.research.google.com/drive/1_mNak1RCxehWBn8Le3LGGHCfbwA2xUTg?usp=sharing.

(a) [2 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) [2 pts] Now consider this following filter:

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

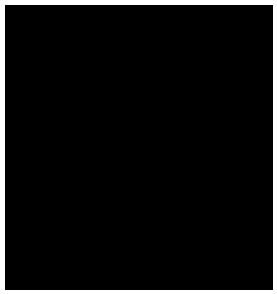
If we apply X as a convolution filter to the original image, which of the following transformed image will we see? **Briefly justify your choice using one or two sentences.**

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

(c) [2 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. 
- B. 
- C. 
- D. 

4. [CNNs] (8 pts) Consider the following 3-layer neural network, with layers enumerated starting from the input.

   (a) **Conv2d**: In channels: 3, Out channels: 5, kernel: 5x5, Stride 1, Padding 0

   (b) **ReLU**

   (c) **MaxPool2d**: kernel: 2x2, Stride 2, Padding 0

   (d) **Conv2d**: In channels: 5, Out channels: 10, kernel: 3x3, Stride 1, Padding 0

   (e) **ReLU**

   (f) **MaxPool2d**: kernel: 2x2, Stride 2, Padding 0

   (g) **Conv2d**: In channels: 10, Out channels: 20, kernel: 3x3, Stride 1, Padding 0

   (h) **ReLU**

   (i) **MaxPool2d**: kernel: 2x2, Stride 2, Padding 0

   If the input image is of size 232 (dimension 1) x 232 (dimension 2) x 3 (dimension 3), compute the following:

   (a) Output size dimension 1

   (b) Output size dimension 2

(c) Output size dimension 3

(d) Number of learnable parameters

2 Python Programming Questions

A IPython notebook is linked on the class website. It will tell you everything you need to do, and provide starter code. Remember to include the plots and answer the questions in your written homework submission!