Problem 1. (PCA) 4 points

Consider the following **standardized** (i.e., centered and scaled by standard deviation) dataset $X \in \mathbb{R}^{3\times 2}$ with 3 data entries and two features, where each data point is labeled as Class A or B:

$$X = \begin{bmatrix} x_1^1 & x_2^1 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 1 & -1 \end{bmatrix} \qquad Y = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix} = \begin{bmatrix} A \\ B \\ A \end{bmatrix}$$

1. What is the dimension of the covariance matrix for the dataset X?
2. The eigenvalues of $X^\top X$ are $\lambda_1 = 4, \lambda_2 = 2$ and the corresponding eigenvectors are $v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, v_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$. Project the data onto the first principal component.
3. Classify the new data point $x^4 = \begin{bmatrix} -1, 0 \end{bmatrix}$ based on its nearest neighbor with the **projected features**. Note that x^4 is already standardized.

Solution:

1. The dimension of the covariance matrix for the dataset X is 2×2.
2. The data is projected onto v_1:

$$Xv_1 = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ -\sqrt{2} \\ 0 \end{bmatrix}.$$

3. Project the data $x^4 v_1 = \begin{bmatrix} -1, 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = -\frac{\sqrt{2}}{2}$. Since $x^4 v_1$ is equally close to $x^2 v_1$ and $x^3 v_1$, the data point x^4 has an equal probability of belonging to class A or B

   (Note: students receive full marks whether they classify the point to class A or B).

Problem 2. (Neural networks) 4 points

Consider a neural network

$$f : [-2, 2] \to \mathbb{R}, \quad \text{with} \quad f(x) = W^{[1]T} \, g\left(W^{[0]T} x + b^{[0]}\right) + b^{[1]}$$

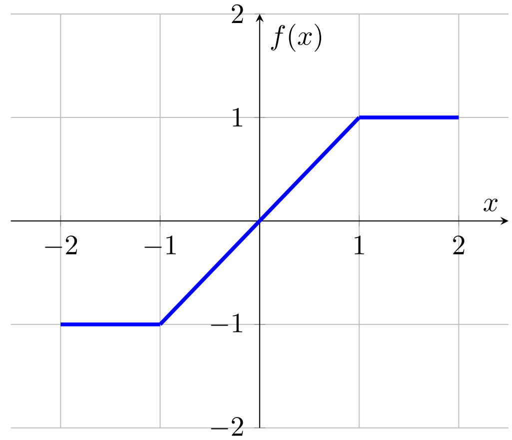
with a single hidden layer and the ReLU activation function $g(x) := \max(0, x)$. It is given that $W^{[0]} = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $b^{[0]} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

1. (1 point) Write the expression for $g\left(W^{[0]T} x + b^{[0]}\right) \in \mathbb{R}^2$ as a function of $x \in \mathbb{R}$.
2. (3 points) Determine $W^{[1]} \in \mathbb{R}^{2\times 1}$ and $b^{[1]} \in \mathbb{R}$ such that f has the graph below.

Solution:

(1) $g\left(W^{[0]T}x + b^{[0]}\right) = \begin{bmatrix}\max(0, x+1) \\ \max(0, x-1)\end{bmatrix}.$

(2) $W^{[1]} = \begin{bmatrix}1 \\ -1\end{bmatrix}$, $b^{[1]} = -1.$



Problem 3. (K-means clustering) 2 points

You are given a data set $X \in \mathbb{R}^{10\times 2}$ for which you have used k-means clustering with $k = 2$ to cluster your data. The center of each cluster is $\mu^1 = (3, 2)$ and $\mu^2 = (7, 8)$. Consider a sample $x \in \mathbb{R}^2$ with a missing value, namely $x = (?,\ 4)$.

1. (1 point) To which cluster center this point is closest?
2. (1 point) Determine the missing entry based on the k-means clusters.

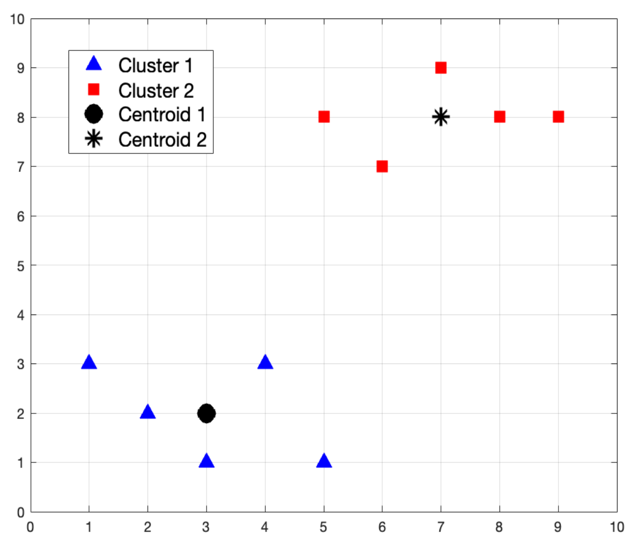


Figure 1: Plot of the samples from the data set X excluding x^{11}

Solution:

1. To find which cluster x belongs we first calculate the Euclidean distance between x and the cluster means μ^1 and μ^2 using only the available components of x. The distances are:

$$f^1 = f(x, \mu^1) = |4 - 2| = 2$$
$$f^2 = f(x, \mu^2) = |4 - 8| = 4$$

Thus, x is closer to the center of cluster 1.

2. To impute the value of x_1, we simply take the value of $\mu_1^1 = 3$. So, the imputed value for x is $(3, 4)$.