Problem 1: (20%) Suppose we have a collected the following one dimensional samples from two classes: $D_1 = \{-3,-2,1\}$, $D_2 = \{0,5,6,7\}$

a) (10%) Draw the decision regions and decision boundaries for nearest neighbor approach with $k=3$, that is 3NN

b) (10%) Can you remove any sample from your data and still have the same decision boundaries?

Problem 2: (10%)

Suppose we have 2 three dimensional training samples from class 1: $\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$

and 2 three dimensional training samples from class 2: $\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$.

Write down a linear decision function $g(x)$ which is positive for all samples from class 1 and negative for all samples from class 2. You do not have to “run” Perceptron algorithm for this problem, you can do it by observation. Of course, if you do wish to “run” Perceptron algorithm, you can.
Problem 3: Suppose we have 2 two dimensional training samples from class 1: \[
\begin{bmatrix}
2 \\
2
\end{bmatrix}, \begin{bmatrix}
2 \\
1
\end{bmatrix}
\]
and 3 two dimensional training samples from class 2: \[
\begin{bmatrix}
2 \\
-1
\end{bmatrix}, \begin{bmatrix}
0 \\
0
\end{bmatrix}, \begin{bmatrix}
2 \\
0
\end{bmatrix}
\]

(a) (10%) What is the weight vector \(a\) after you apply batch Perceptron algorithm for one iteration, starting with weight vector \(a = \begin{bmatrix}
1 \\
-1 \\
-1
\end{bmatrix}\) and using step size \(\eta = 2\).

(b) (10%) What is the value of the Perceptron objective function for the weight vector \(a\) computed in part (a)?

(c) (15%) Plot the samples, the decision boundary corresponding to the weight vector \(a\) computed in part (a), and also the normal \(w\). Make sure to label your axis.
Problem 4:

(a) (10%) Write down the discriminant function corresponding to class 2 in the neural network above (the function corresponding to output unit 2). Assume that for the hidden unit and the output units we have the same function $f$.

(b) (15%) Suppose $f(x)$ is the standard thresholding function ($f(x) = 1$ for positive $x$ and $f(x) = -1$ otherwise). Suppose you input example $x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ to the network above. Is it going to be classified as class 1 or class 2? Recall that you assign the class corresponding to the higher value of discriminant function.
Problem 5 (10 %): Suppose we have collected the following one dimensional samples from two classes: D1={-1,1}, D2={0,4,5,6}, and you use 1NN (nearest neighbor with k = 1) classifier. What is the 3-fold cross validation error? You may wish to draw the data on the plot below, but you don’t have to. **Show all the work!**