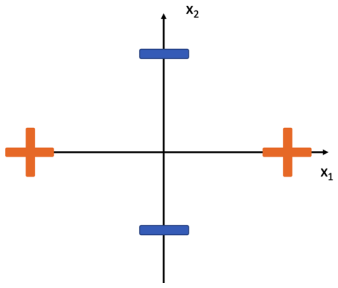
Computer Science Department
California State University Channel Islands

COMP 478 - Homework 3

Deadline: 4/26/2022, 11:59 am

1. (15 points) Consider the dataset below where $x_{1,2}$ = {(2, 0),(-2, 0)} belong to the + class and $x_{3,4}$ = {(0, -2),(0, 2)} belong to the - class. We want to employ a polynomial kernel (with degree 2) to make the data linearly separable.



   (a) Consider our original feature space is X = [X_1, X_2]^T . Find the new transformed feature space.

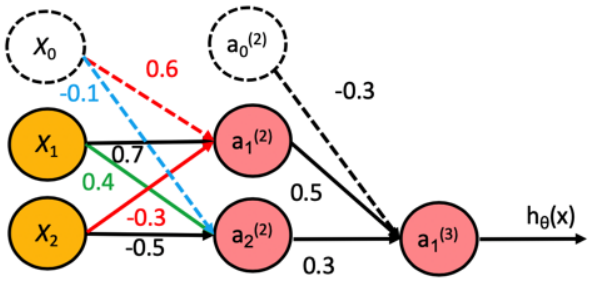
   (b) Plot the transformed data in a 2d graph (pick only two new features to have linearly separable data). Find the maximum margin hyperplane in this space.

   (c) Plot the maximum margin hyperplane in the original feature space.

2. (5 points) In Lecture 15, slide 10, find $\phi(x_i)$ if:

$$\text{kernel} = K(x_i,x_j) = (<x_i,x_j>)^3$$

3. (5 points) Consider the following neural network with all the weights for the edges:



consider input values $X_1 = 1$, $X_2 = 1$, and rectified linear function (ReLU) as the activation function for all the nodes. Find the activation values for ${a_1}^{(2)}$,${a_2}^{(2)}$, and ${a_1}^{(3)}$.

4. (35 points) You want to train the following model with a single data point ($x_0 = 1, y_0 = 90$) using SGD:

$$z = w_0 + w_1 x + w_2 x^4$$
$$y = 100z + e^z$$
$$Loss = L = (y - y_0)^2$$

   (a) Draw the computational graph for the given model.

   (b) Run the forward pass to calculate the values of all the edges. ($w_0 = -1, w_1 = 1, w_2 = 1$)

   (c) Calculate the gradient terms for your computational graph using backprop algorithm. ($dL/dw_0, dL/dw_1, dL/dw_2, dL/dx$, ...)

   (d) If the learning rate = 0.001, what are the updated values for w_0, w_1, w_2?

5. (20 points) We want to use a Convolutional Neural Network (CNN) for a classification task with 4 classes. The input is a 25x25 gray-scale image. Our CNN has the following layers from input to output: a convolution with 3 filters (5x5, stride = 1) and ReLU activation, a max pooling layer (filter size: 3x3, stride = 1) , and finally a fully-connected layer. Please answer the following questions about this network.

   (a) How many parameters in the convolutional layer do we need to learn?

   (b) How many parameters do we need to learn for the entire network?

   (c) (True or False-and explain why?) A neural network with FC layers (the same size as our CNN) can outperform the above convolutional network.

   (d) What is the advantage of a CNN compared to a fully-connected neural network with the same size layers?

6. (20 points) The input of a convolutional layer with one filter (size = 2 × 2 , stride =2) is a grey-scale image shown on the left.

170	50	20	20
1	10	0	11
200	90	0	0
17	10	35	0
100	10	5	1
100	10	23	0

71	-54
61	1
41	5

   (a) If the output of the convolutional layer is the image on the right, find the parameters of the filter.

(b) Find the resulted image if we use zero-padding (p=1).

(c) If we use our resulted image from convolution layer (in part b) as the input of a pooling layer with a filer (size = 2×2 , stride =1), what is the output image?