Classification question

a) Assume the following training data in the two-dimensional plane of X_1 and X_2 is available (Figure 1). The target variables for the points in the red and blue are +1 and -1. We summarise the data as the following tuples: <(2,0), 1>, <(0,2),-1>, <(0,-2),1>, and <(-2,0),1>, respectively.

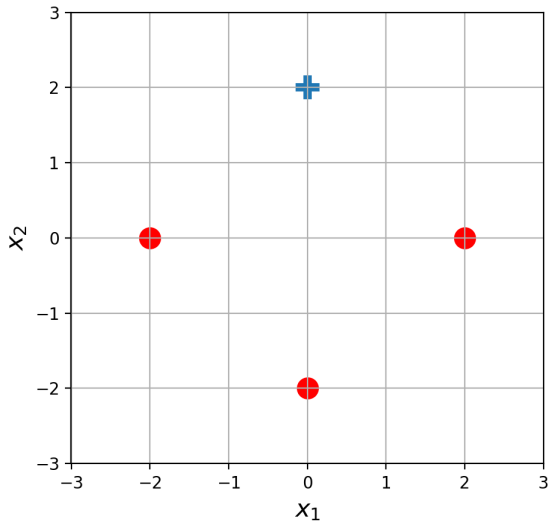

Figure 1

i. Design a k-NN classifier with k=1 and use it to determine the class variables C_1 through C_4 for the following test data points: <(0,1), C_1>, <(1.5,1), C_2 >, <(-0.5,1), C_3 >, and <(0,0), C_4 >:

[4 marks]

SOLUTION:
C_1 = -1, C_2 = 1, C_3 = -1, C_4=unknown

ii. What would be the class variable C_4 above if we had used k=3? [2 marks]
SOLUTION:
C_4 = 1

iii. Write down the equations that specify the decision boundary between the two classes. [4 marks]
SOLUTION:
X_1 – X_2 = 0 in X_1 > 0.0 and X_2 > 0.0 2 marks
X_1 + X_2 = 0 in X_2 > 0.0 and X_1 < 0.0 2 marks

b) In the same data set in Figure 1, we apply a linear SVM model with the predictor $y(X_1, X_2)$ for classification.

I. Which data points are the support vectors? Write down the equation for $y(X_1, X_2)$. (Hint: First visually assess the data to determine the decision boundary and the support vectors. Observe the constraints for the margin and SVM classifier.)

[6 marks]

SOLUTION:

(2,0), (0,2), and (-2,0) 3 mark,

$y(X_1, X_2) = -X_2 + 1$ 3 marks

II. Specify the Lagrange multipliers $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ for each of the data points in the training data (2,0), (0,2), (-2,0), and (0,-2), respectively.

[4 marks]

SOLUTION:

$\alpha_1 = .25, \alpha_2 = .5, \alpha_3 = 0.25, \alpha_4 = 0$