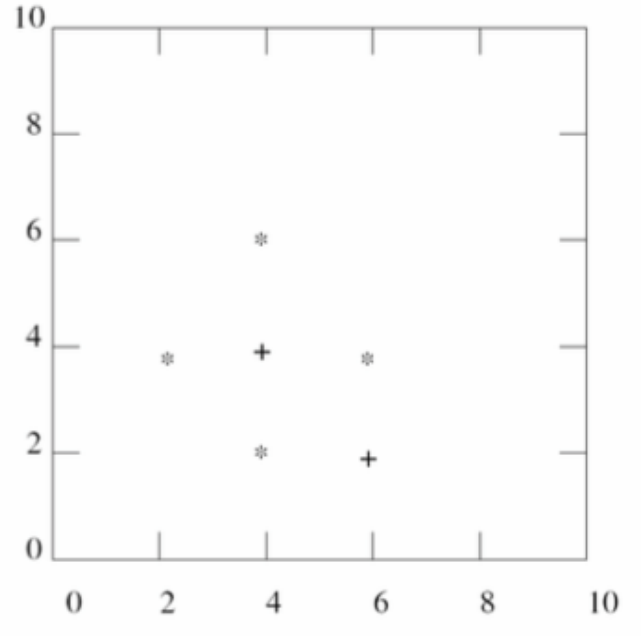
INSTANCE BASED LEARNING

1. Given the following data items and class labels for a 1-D scenario.

Data Item	x_1	Class Label
1	0.0	+1
2	1.5	+1
3	2.0	-1
4	2.8	-1
5	3.9	-1

You decide to perform 5-fold cross-validation on this dataset using 1-NN. What is the value of total error? Repeat the process with 3-NN and calculate the error?

2. In the following questions you will consider a *k*-nearest neighbor classifier using Euclidean distance metric on a binary classification task. We assign the class of the test point to be the class of the majority of the *k* nearest neighbors.



a. Sketch the decision boundary of 1-NN classification

b. If you perform 6-fold cross validation using 1-NN, what will be the total training error?

c. How would the point (8,1) be classified using this dataset?

3.

1. Consider the following training set in the 2-dimensional Euclidean space:

x	y	Class
−1	1	−
0	1	+
0	2	−
1	−1	−
1	0	+
1	2	+
2	2	−
2	3	+

Figure 1 shows a visualization of the data.

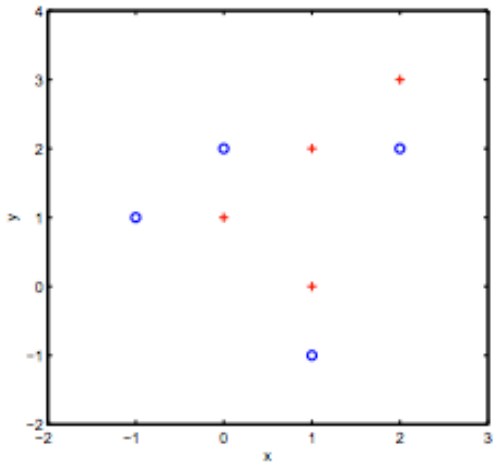


Figure 1: Dataset for Problem 2

(a) (1pt) What is the prediction of the 3-nearest-neighbor classifier at the point (1,1)?

(b) (1pt) What is the prediction of the 5-nearest-neighbor classifier at the point (1,1)?

(c) (1pt) What is the prediction of the 7-nearest-neighbor classifier at the point (1,1)?

4. For leave-one-out cross validation approach, can we say that the errors for each of the trials are independent of each other? Explain

5. Can you think of a scenario where:
a. the 1-NN has lower training error than SVM
b. the 1-NN has higher training error than SVM