

Machine Learning

V. Adamchik

CSCI 567

Spring 2019

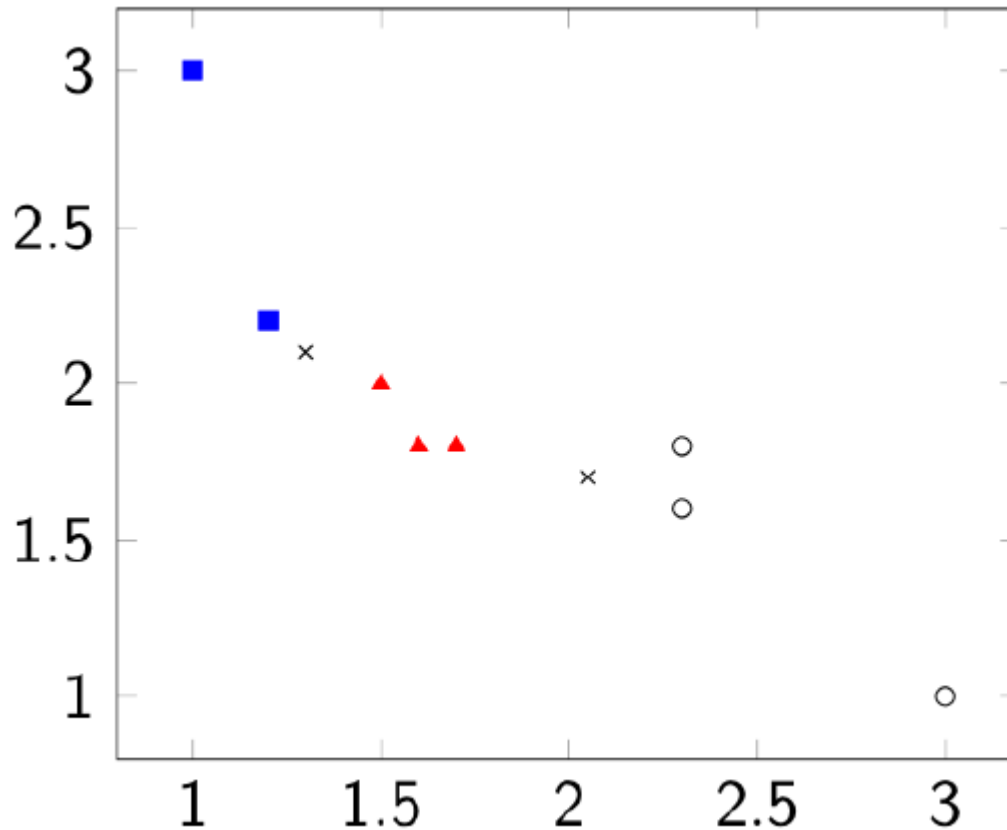
Discussion 1

University of Southern California

k-NNC

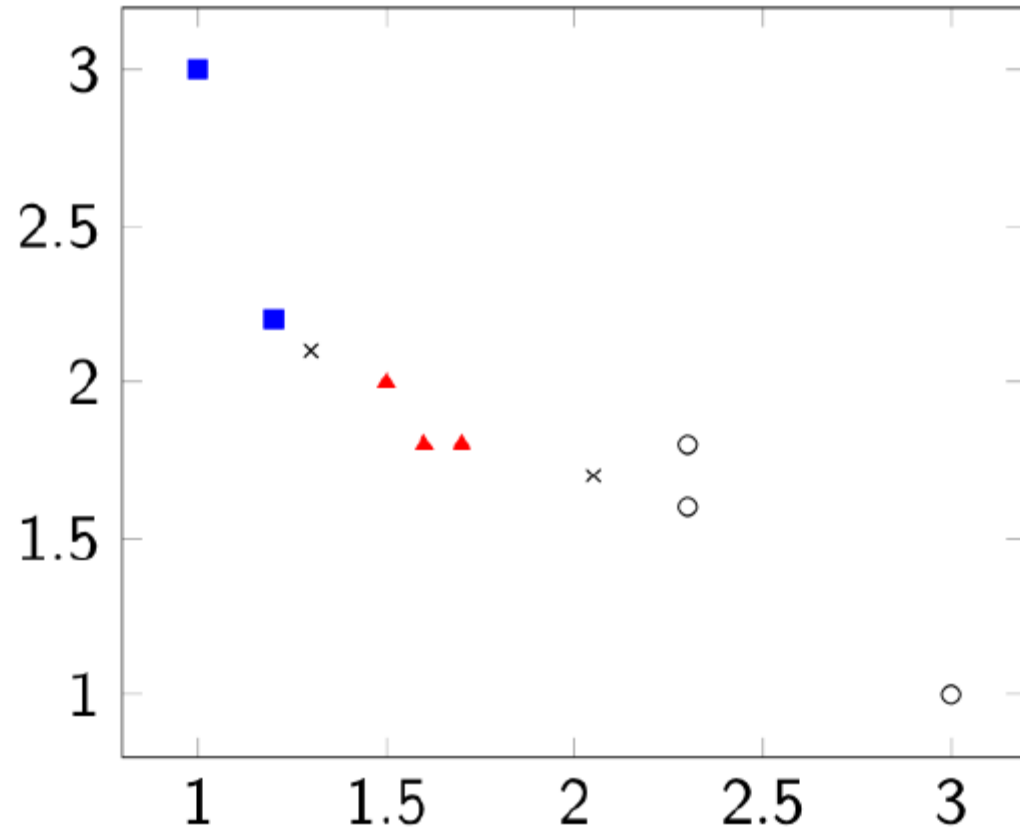
Problem 1a)

Consider the following data; triangles, squares, and circles are three classes of data in the training set; the x items are the unlabeled test data.



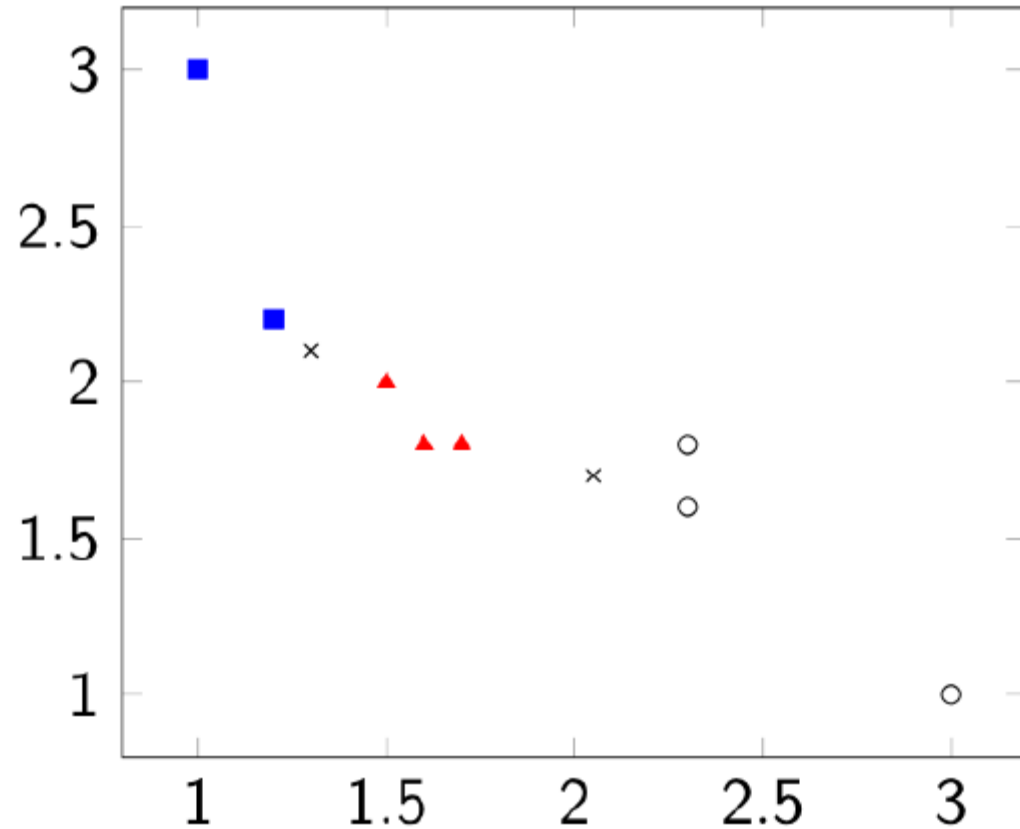
What is the test-data label x be if $k = 1$?

Problem 1b)



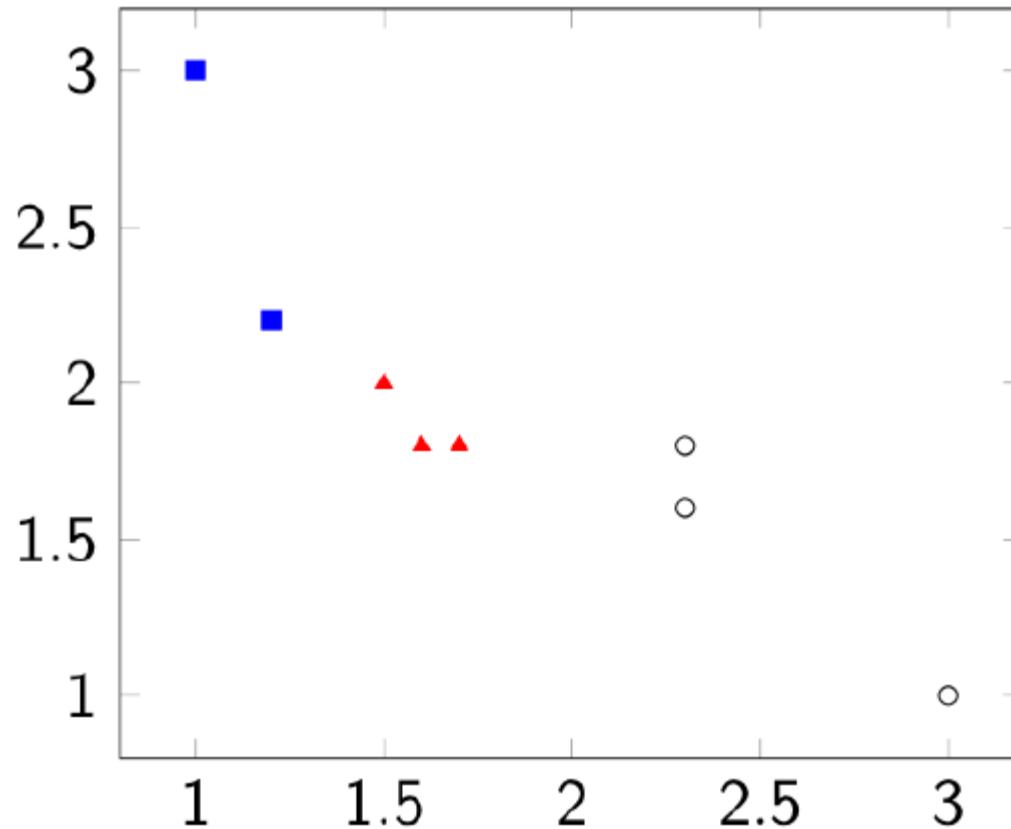
What is the test-data label x be if $k = 3$?

Problem 1c)



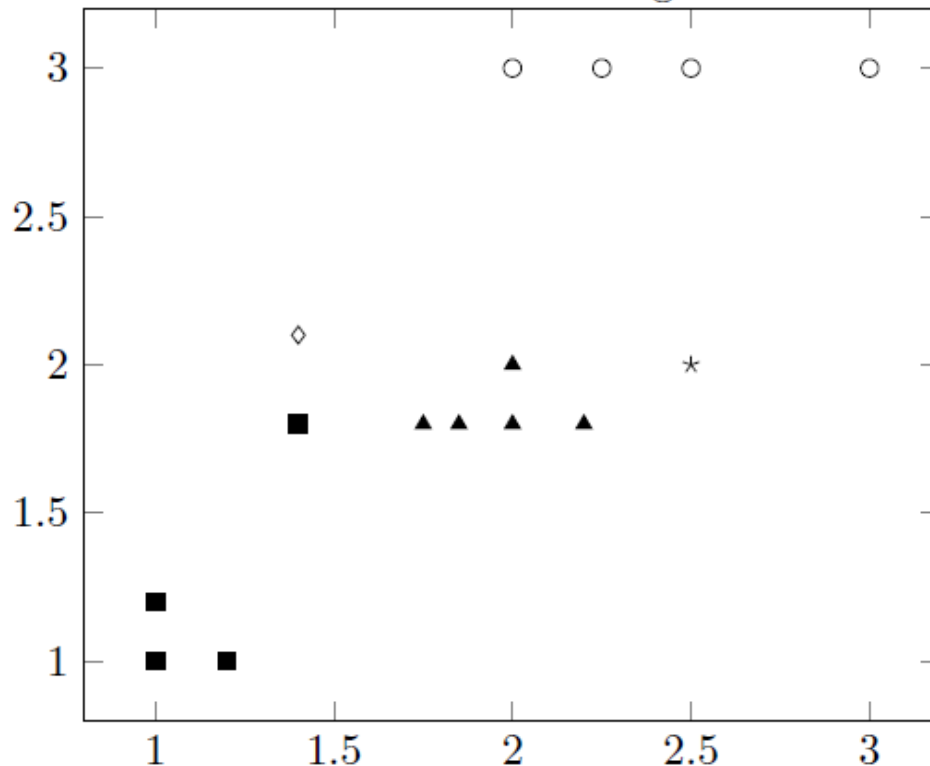
What is the test-data label x be if $k = 5$?

Problem 2



What is the accuracy and error with leave-one-out, assuming $k = 1$?

Problem 3

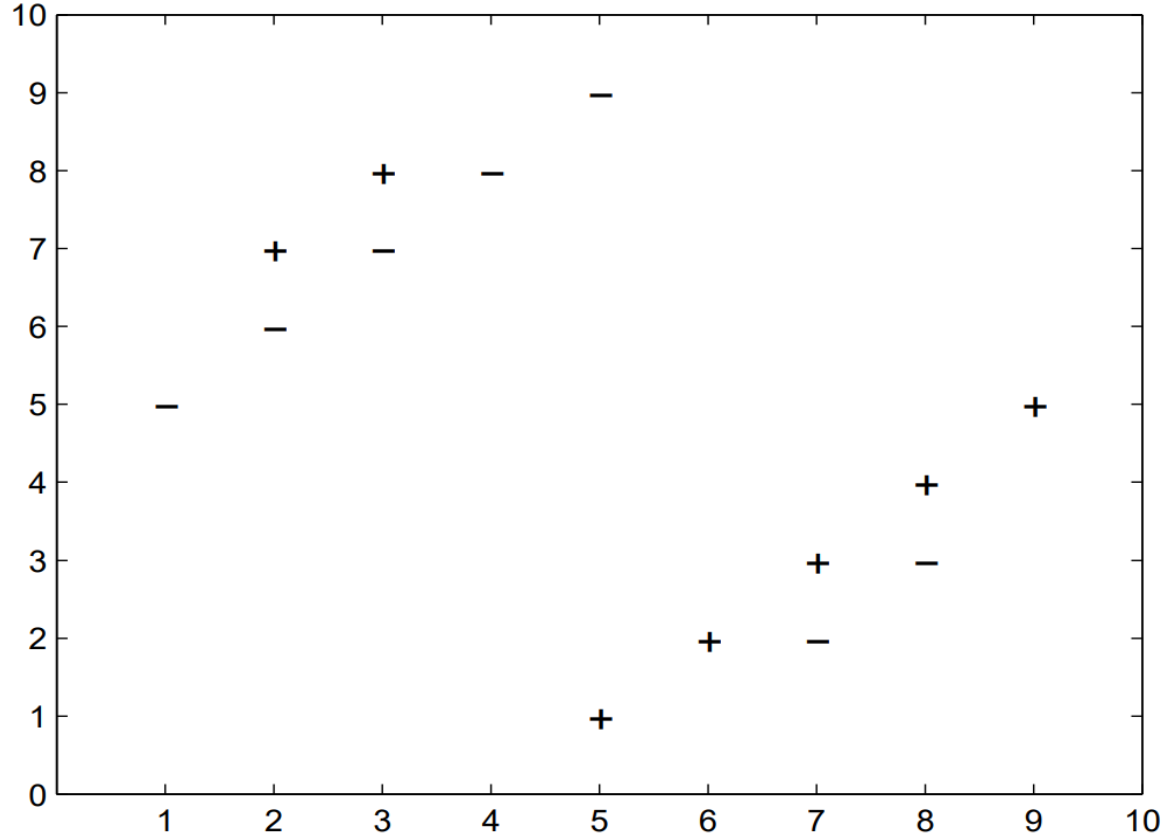


How many training data points will be misclassified with leave-one-out ($k = 1$)?

What is the smallest value of k to always classify the diamond as class triangle?

What is the smallest value of k to classify the star as open circle?

Problem 4



What value of k minimizes leave-one-out cross-validation error for this dataset?

ML Concepts

What is a *parametric* model?

Which of the following phenomenon is called *overfitting*?

(A) low training error, low test error

(B) low training error, high test error

(C) high training error, low test error

(D) high training error, high test error

(T/F) k-NN it is very likely to overfit due to the curse of dimensionality.

What is the popular solution for overfitting?

ML Concepts

What is a S -fold *cross-validation* in your own words?

How do you partition the training data into S sets?

For small datasets, what is a good choice for S ?

Is there is an a priori good choice of S ? What is the rule of thumb choice of S ?

ML Concepts

Which of the following is the Bayes optimal classifier?

- (A) A randomized classifier that predicts y with probability $P(y|x)$ given x .
- (B) A deterministic classifier that predicts $\operatorname{argmax}_y P(y|x)$ given x .
- (C) A deterministic classifier that predicts $\operatorname{argmin}_y P(y|x)$ given x .
- (D) Both (A) and (B).

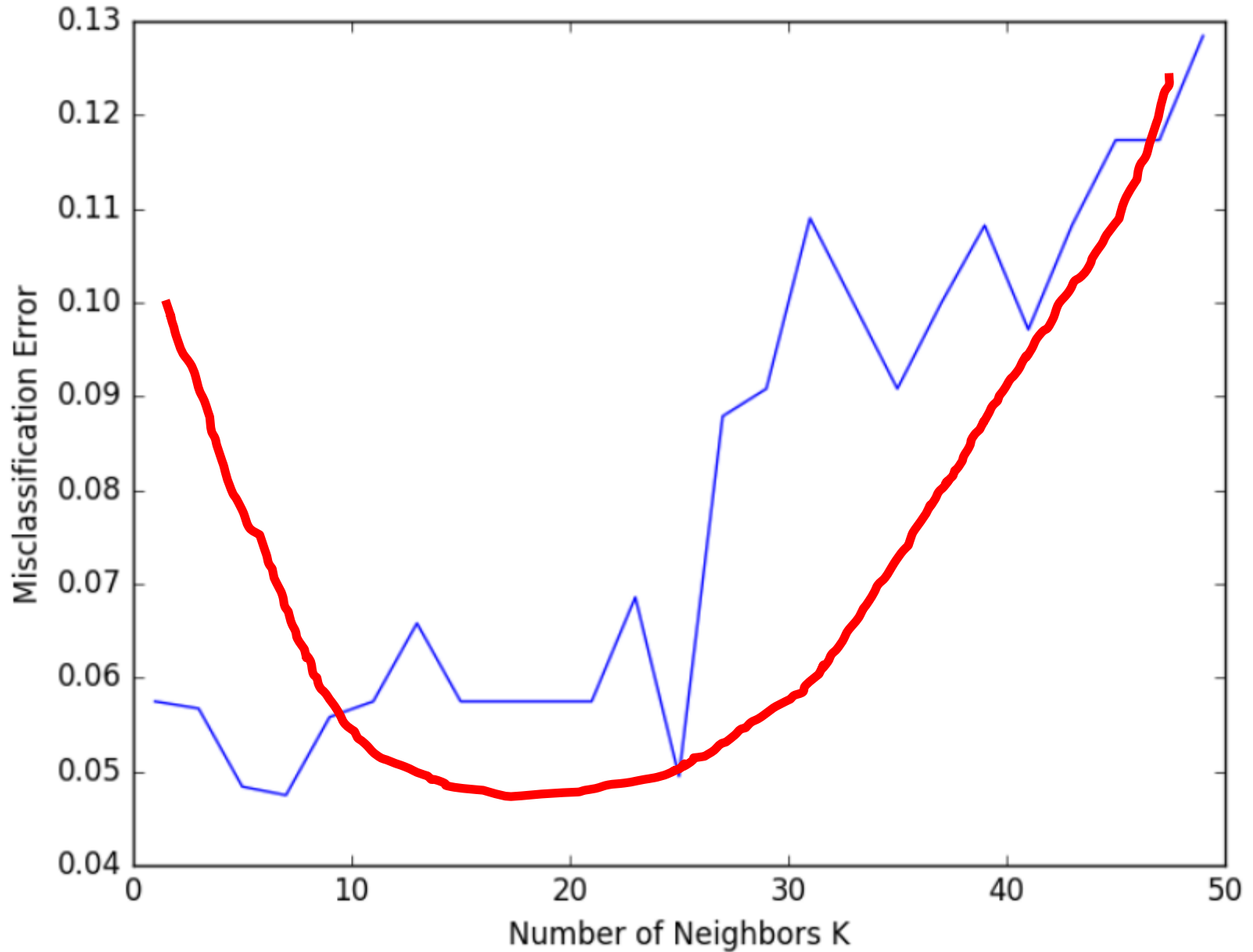
Hyperparameter k

The k -nearest neighbor classifier requires a setting for k .
But what number works best?

Can we use the training data?

How do you find k if there is no validation set?

Hyperparameter k



Digit Recognition

7291 train points, 2007 test points

Error rates:

Neural net: 0.049

1-NN/Euclidean distance: 0.055

1-NN/tangent distance: 0.026

