1

1. Since we are predicting the value of a continuous/numeric variable (CEO salary), this problem is a regression problem.

number of data points (n) = 500 (top 500 firms in the US)

number of features (p) = 3 (profit, number of employees, industry)

1. Since we are predicting the value of a categorical /binary variable (success or failure), this problem is a classification problem.

number of data points (n) = 20 (20 similar previously launched products)

number of features (p) = 13 (price charged for the product, marketing budget, competition price, and ten other variables)

1. Since we are predicting the value of a continuous/numeric variable (% change in the US dollar), this problem is a regression problem.

number of data points (n) = 52 (Number of weeks in 2012)

number of features (p) = 3 (% change in the US market, % change in the British market, and the % change in the German market)

2.

(a) for observation 1, = 3

for observation 2, = 2

for observation 3, = 3.1623

for observation 4, = 2.23607

for observation 5, = 1.4142

for observation 6, = 1.73205

(b) For the given test point, the nearest training point is observation 5 (based on the calculated Euclidean distances in the previous problem). Therefore, for K=1, the prediction for the test point is “Green” (which is the prediction for observation 5)

(c) For the given test point, the three nearest training points are observation 5, observation 6, observation 2. Therefore, for K=3, the prediction for the test point is argmax (y belongs to Y) ( 1 (for y= Green), 2 (for y= Red)) = “Red”

3.

(a) 0.1 (As the distribution of X is uniform)

(b) 0.1 x 0.1 = 0.01 (In this equation, one 0.1 is from X1 and other 0.1 is from X2)

(c) 0.1^(100) (Here, 0.1 fraction from each of the 100 features are multiplied by each other)

(d) When there is only one feature (in problem a), we are using 10% (0.1) of the total observations, but when we started increasing the number of features, the total available observations for test prediction started decreasing as we have more constraints on the data to use from the new features that are being added. As it can be seen, with the two features, the fraction of available observations for test prediction is reduced to 0.01 (problem b), and with 100 features, it is reduced further to 0.1^100 (problem c). To generalize, for n features, in this case, it would be 0.1^n.

(e)As the p-dimensional hypercube (with a side length a) contains 10% of the training data, volume of the hyper cube should be 0.1 (or) 10% of the available training data space

4

1. Accuracy = TP+TN/(TP+FN+FP+TN) = (8+974)/(8+2+16+974) = 0.982
2. Precision = TP/(TP+FP) = 8/(8+16) = 0.3333
3. Recall = TP/(TP+FN) = 8/(8+2) = 0.8

5

|  |  |
| --- | --- |
| **Confidence positive** | **Correct class** |
| 0.95 | + |
| 0.85 | + |
| 0.8 | - |
| 0.7 | + |
| 0.55 | + |
| 0.45 | - |
| 0.4 | + |
| 0.3 | + |
| 0.2 | - |
| 0.1 | - |

TPR = 2/6

FPR = 0/4

TPR = 4/6

FPR = 1/4

TPR = 6/6

FPR = 2/4

TPR = 6/6

FPR = 4/4