

Div: B

Roll No.: 422004

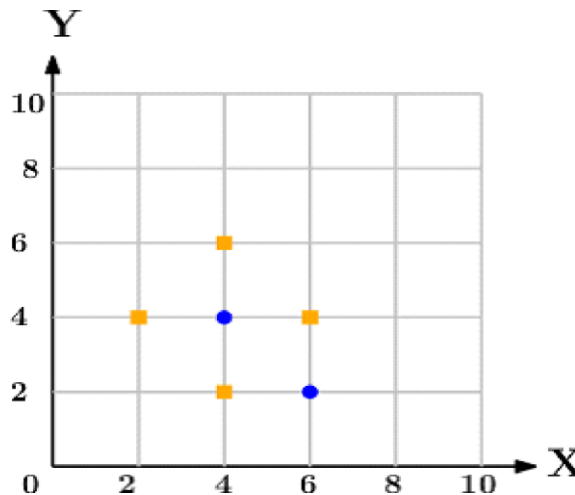
Gr. No.: 21810939

Name of Student: Aakanksha Bhondve

subject: Advanced Machine Learning

Assignment No. 3

Aim: In the following diagram let blue circles indicate positive examples and orange squares indicate negative examples. We want to use k-NN algorithm for classifying the points. If $k=3$, find the class of the point (6,6).



Theory:

1) Classification algorithms:

1. **Logistic Regression:** It is a very basic yet important classification algorithm in machine learning that uses one or more independent variables to determine an outcome. Logistic regression tries to find a best-fitting relationship between the dependent variable and a set of independent variables. The best-fitting line in this algorithm looks like S-shape.

2. **Naive Bayes:** Naive Bayes is based on Bayes's theorem which gives an assumption of independence among predictors. This classifier assumes that the presence of a particular feature in a class is not related to the presence of any other feature/variable.

Naive Bayes Classifier are of three types: Multinomial Naive Bayes, Bernoulli Naive Bayes, Gaussian Naive Bayes.

3. **K-Nearest Neighbour Algorithm:** KNN works on the very same principle. It classifies the new data points depending upon the class of the majority of data points amongst the K neighbour, where K is the number of neighbours to be considered. KNN captures the idea of similarity (sometimes called distance,

proximity, or closeness) with some basic mathematical distance formulas like euclidean distance, Manhattan distance, etc.

4. SVM: SVM stands for Support Vector Machine. This is a supervised machine learning algorithm that is very often used for both classification and regression challenges. However, it is mostly used in classification problems. The basic concept of the Support Vector Machine and how it works can be best understood by this simple example. So, just imagine you have two tags: green and *blue*, and our data has two features: x and y . We want a classifier that, given a pair of (x,y) coordinates, outputs if it's either *green* or *blue*. Plot labeled training data on a plane and then try to find a plane (hyperplane of dimensions increases) that segregates data points of both colors very clearly.

5. Decision Tree: The decision tree is one of the most popular machine learning algorithms used. They are used for both classification and regression problems. Decision trees mimic human-level thinking so it's so simple to understand the data and make some good intuitions and interpretations. They actually make you see the logic for the data to interpret. Decision trees are not like black-box algorithms like SVM, Neural Networks, etc.

2) Working of K-NN algorithm:

The K-NN working can be explained on the basis of the below algorithm:

Step-1: Select the number K of the neighbours

Step-2: Calculate the Euclidean distance of K number of neighbours

Step-3: Take the K nearest neighbours as per the calculated Euclidean distance.

Step-4: Among these k neighbours, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbour is maximum.

Step-6: Our model is ready.

Implementation link:

https://colab.research.google.com/drive/1BUHnDu_SV9inaK6vkTQEibnzb9-vs2Bm#scrollTo=31CxNgahXdnA

Code & Output:

