Machine Learning Assignment I

S20170010016 – Atul Vidyarthi
S20170010073 – Kiran Sankar E J

KNN Classifier

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).

We use Minkowski distance as the similarity measure. The optimal k and p values are found out through 3 – fold cross validation. We consider one out of three random segments of the training set to be the validation set and use it to find the k and p values for which the error is minimum.

KNN classifier predicts the class of a new object by comparing its features with its k-nearest neighbours in the feature space. This method of classification provides easy to interpret output, comparatively better calculation time and predictive power.

The problem that may occur here is to find out the optimal k value, which is done using r-fold cross validation, in this case r is 3. We also have to find out the best possible p for the Minkowski distance.

The accuracy is calculated as the ratio of the number of matching predictions to that of the total number of records in the validation list.