Birla Institute of Technology and Science, Pilani

BITS-F464: Machine Learning
2^{nd} Semester 2018-19

Labsheet-05: Rote, CBR, K-NN

In this lab we would implement three methods for classification *viz*. Rote, Case-Based Learner (CBR) and K-NN.

1 Rote learner

It is a memorization technique. The classifier searches in its stored database for the answer. If found it returns the label of retrieved item otherwise, flags a message saying "unable to classify".

2 Case-based reasoning (CBR)

In CBR, the training examples, the cases, are stored and accessed to solve a new problem. To get a prediction for a new example, those cases that are similar, or close to, the new example are used to predict the value of the target features of the new example. This is at one extreme of the learning problem where, unlike decision trees and neural networks, relatively little work must be done offline, and virtually all of the work is performed at query time. Case-based reasoning is used for classification and for regression. It is also applicable when the cases are complicated, such as in legal cases, where the cases are complex legal rulings, and in planning, where the cases are previous solutions to complex problems. In our case assume CBR outputs the classification based on the nearest data point only.

3 k-NN

k-nearest neighbors algorithm (k-NN) is a non-parametric classification method. It first determined k closest training examples in the feature space and then classification output depends on the plurality vote of its neighborhood. Object is assigned the class most common among its k nearest neighbors. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. See following code

```
data(iris)

## Split in train + test set
idxs <- sample(1:nrow(iris),as.integer(0.7*nrow(iris)))
trainIris <- iris[idxs,]
testIris <- iris[-idxs,]
```

```
## A 3-nearest neighbours model with no normalization
nn3 <- kNN(Species ~ ., trainIris, testIris, norm=FALSE, k=3)

## The resulting confusion matrix
table(testIris[,'Species'], nn3)

## Now a 5-nearest neighbours model with normalization
nn5 <- kNN(Species ~ ., trainIris, testIris, norm=TRUE, k=5)

## The resulting confusion matrix
table(testIris[,'Species'], nn5)
```

4 Questions

Use `iris` dataset that is inbuilt in r. It has three species (setosa, versicolor, virginica). Assume complete database is given for learning. For every item in the database

1. Determine its label with rote learner and report the accuracy accuracy.
2. Add random noise in each of the four values (Sepal.Length Sepal.Width Petal.Length Petal.Width). Now apply CBR to find the classification. Report accuracy.
3. In the above setting (random noise added to values) apply k-NN to determine accuracy for k = 2, 3, 4, 5, 6, 7, 8, 9, 10