

Locally Adaptive Nearest Neighbours

Blue Octopus (Team 46)

- Aparna Agrawal
- Aditya Sehgal
- Akshit Gureja
- Tanmay Goyal





K-Nearest Neighbour

- Stores entire training set in memory
- Test sample is labelled based on the majority of k-nearest neighbours.
- 'K' is global and determined via cross-validation

Limitations (Motivation of Research Paper)

- 'K' is not adaptive to characteristics like noise and irrelevant features.
- The pattern of the distribution might vary in the feature space.
- For certain applications, locally adaptive 'k' has much better accuracy because the 'k' is unique to the surrounding data.



Locally Adaptive Nearest Neighbour

- Comparable in computational complexity and accuracy to k-NN
- Can be constructed at classification time.
- Eliminates need for global cross-validation to get correct 'k'

Approach:

There are 2 approaches to compute locally varying k:

1. Single 'k' or set of values of 'k' for each training pattern
2. Find k-values for training patterns after combining them into groups.

Methods Compared:

- **K-nearest neighbour (KNN)**
- **Unrestricted local KNN**
- **Pruned local KNN**
- **One k-per class local KNN**
- **One k-per cluster local KNN**

K-Nearest Neighbour



- Most fundamental ML model
- Stores all the training examples
- Single global value of 'k'
- determined during training
- Testing data is classified based on the majority of these k-nearest neighbours.

Unrestricted Local KNN

1 Store the training data set

2 Get list of 'k' for every training sample based on leave-one-out cross validation

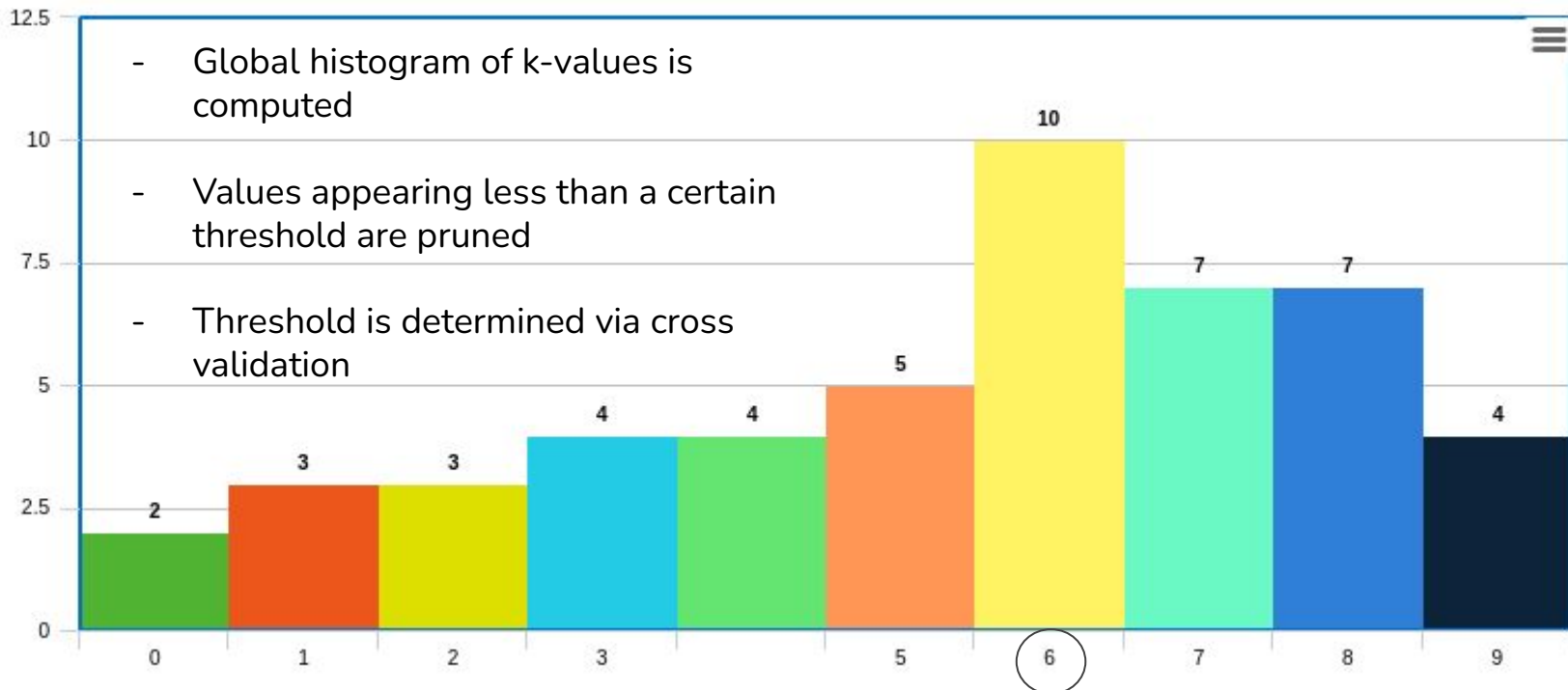
3 Get the M-nearest neighbours for the test sample 'q'

4 Get the 'k' which classifies the majority of these M neighbours as $K_{M,q}$

5 Use this $K_{M,q}$ to classify the test sample

Pruned local KNN

- Global histogram of k-values is computed
- Values appearing less than a certain threshold are pruned
- Threshold is determined via cross validation



One k-per class local KNN

STEP 1

For each output class,
determine value of 'k' that
would correctly classify
maximum number of training
samples from that particular
class.

STEP 2

To classify a test query 'q'

The query is assigned to
class C_i , for which the
percentage of k_i nearest
neighbours of 'q',
belonging to class C_i is
maximum.



One k-per cluster local KNN

Step 1

Use an unsupervised clustering algorithm to determine the clusters of the training data.

Step 2

A single value of 'k' is determined for each cluster obtained above.

Step 3

Classify each query according to the 'k' value of the cluster it is assigned to.

Thank You