MMK
Lehrstuhl für
Mensch-Maschine-Kommunikation

Pattern Recognition

Tutorial No. 5
16-17/06/2016

Exercise 12 – kNN classification speedup

The speed of kNN classifiers is greatly affected by the size of the data set: to classify a sample you have to compute the distances to all of the training set vectors. To try and speed it up by choosing representative samples of each class instead of using all the samples as reference vectors is the main objective of this exercise.

You are provided with an implementation of Mean-Shift vector quantization. The output of the algorithm provides you with a codebook of representative vectors of the underlying distribution.

Your task is to retrieve a reduced number of reference vectors for each class (lines 38-39): $\{W^i_{\Omega_c}\}_{i=1}^{M_c}$ for each class $\Omega_c \in \{0, 1\}$.

Then implement the following steps of the kNN algorithm in `knn_fit_predict` function (for classifying the vector x):

a) Calculate the distances between x and the reference vectors (line 86)

b) Sort the reference vector class labels by the distance to the classified data point (line 92 already implemented)

c) Assign a class label to x based on the k closest reference vectors (line 99)

What is the theoretical speedup for performing kNN classification with a total of $M = \sum_{\Omega_c \in \{0,1\}} M_c$ reference vectors as training data instead of a total of N data points in the training set?

Observe the speedup and quality of the kNN classifier using the reduced number of representative vectors. Compare to the theoretical speedup.

Exercise 13 – Decision border and distance metrics

For a two class problem the two reference vectors are $W_{\Omega_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $W_{\Omega_2} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$.

a) Graphically derive the decision border for a Nearest Neighbour (NN) classifier with the Euclidean distance.

b) Mathematically derive the decision border for an NN classifier with the Euclidean distance

c) Now the first class is represented with the two reference vector $W^1_{\Omega_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $W^2_{\Omega_1} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$, and the second class is represented with one reference vector $W_{\Omega_2} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Derive the decision border graphically for an NN classifier

d) Graphically analyze the effect of different covariance matrices for an NN classifier with the Mahalanobis distance. Try out the same covariance matrix for both classes, different covariance matrices for the two classes, the identity matrix and diagonal matrices. (This is best done in MATLAB)

Exercise 14 – NN classification with different distance metrics

For a two class problem the two reference vectors are $W_{\Omega_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $W_{\Omega_2} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, the corresponding covariance matrices (and their inverses are) $C_1 = \begin{bmatrix} 1 & 0.5 \\ 2 & 0.7 \end{bmatrix}$, $C_2 = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$, $C_1^{-1} = \begin{bmatrix} -2.3 & 1.7 \\ 6.7 & -3.3 \end{bmatrix}$, and $C_2^{-1} = \begin{bmatrix} -0.4 & 0.6 \\ 0.6 & -0.4 \end{bmatrix}$.

(**UPDATE 2016-06-16: what is wrong with these matrices?**)

The two unknown vectors $y_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $y_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ shall be classified.

a) Classify the two vectors with an NN classifier and the Euclidean distance

b) Classify the patterns with an NN classifier and the Mahalanobis distance