Neighbourhood Components Analysis & Metric Learning

Sam Roweis

Dept. Computer Science

University of Toronto

roweis@toronto.edu

Abstract

Say you want to do K-Nearest Neighbour classification. Besides selecting K, you also have to chose a distance function, in order to define "nearest". I'll talk about a method for *learning* – from the data itself – a distance measure to be used in KNN classification. The learning algorithm, Neighbourhood Components Analysis (NCA) directly maximizes a stochastic variant of the leave-one-out KNN score on the training set. Of course, the resulting classification model is non-parametric, making no assumptions about the shape of the class distributions or the boundaries between them. I will also discuss an variant of the method which is a generalization of Fisher's discriminant and defines a convex optimization problem by trying to collapse all examples in the same class to a single point and trying to push examples in other classes infinitely far away. By approximating the metric with a low rank matrix, these learning algorithms, can also be used to obtain a low-dimensional linear embedding of the original input features allowing that can be used for data visualization and very fast classification in high dimensions.