Neighbourhood Components Analysis

Say you want to do K-Nearest Neighbour classification. Besides selecting K, you also have to choose a distance function, in order to define "nearest". I'll talk about a novel 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. It can also learn a low-dimensional linear embedding of labeled data that can be used for data visualization and very fast classification in high dimensions. Of course, the resulting classification model is non-parametric, making no assumptions about the shape of the class distributions or the boundaries between them.