TU München
Fakultät für Informatik
PD Dr. Rudolph Triebel
John Chiotellis

Machine Learning for Computer Vision
Summer term 2017

June 17, 2017
Topic: Metric Learning, Boosting

Exercise 1: Metric Learning

a) Given a valid metric D_M, is D_M^2 also a metric? Why?

b) Given a matrix X of n data points $x_i \in \mathbb{R}^d$, show how computing the eigen-decomposition of the covariance of X is equivalent to computing the singular value decomposition of X .

c) What is the difference between metric learning and kernel learning? When would you prefer to use one over the other?

d) In Neighborhood Component Analysis, we define a stochastic neighbor selection rule. The probability that a data point j is selected as neighbor of point i is given by:

$$p_{ij} = \frac{\exp\{-||Lx_i - Lx_j||^2\}}{\sum_{k \neq i} \exp\{-||Lx_i - Lx_k||^2\}} \tag{1}$$

namely a softmax over the squared distances to all points in the transformed space. The goal is to maximize

$$f(L) = \sum_i \sum_{j \in C_i} p_{ij} \tag{2}$$

namely the probability that the neighbors that will be selected for each point i will belong to the same class C_i. Can you derive the gradient of $f(L)$?

Exercise 2: Adaboost (Programming)

Download the file 'banknote_auth.zip' available at the course's website. The data are features of banknotes and the labels indicate whether a banknote is forged or not. The dataset is taken from `https://archive.ics.uci.edu/ml/datasets/banknote+authentication` with some duplicate entries removed. Implement the AdaBoost algorithm with decision stumps as weak classifiers.

a) To begin train on 50% of the data with 20 weak classifiers.

b) Generate a plot of the classification error with respect to the number of weak classifiers. What do you observe?

c) Add more weak classifiers. Does the error still change? What's the optimal number of weak classifiers to use?

d) Now keep the number of weak classifiers fixed and try different training/testing set sizes. How does it affect the classification accuracy?