

Machine Learning for Computer Vision Winter term 2017

November 21, 2017
Topic: Metric Learning

Exercise 1: Metric Learning

- Given a valid metric D_M , is D_M^2 also a metric? Why?
- 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 .
- What is the difference between metric learning and kernel learning? When would you prefer to use a kernel method over a metric learning method?
- 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:

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

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

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

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 $f(L)$?

- What is the difference between LDA and NCA?
- The KL-divergence measures the (dis-)similarity of two probability distributions. It is defined as:

$$D_{KL}(p||q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx$$

Is the KL-divergence a metric? Why?