Dimensionality Reduction by Minimum Margin Sandwich

Sreangsu Acharyya Joydeep Ghosh

Abstract

The nonlinear dimensionality reduction problem is posed as finding two parallel hyperplanes in a feature space that sandwiches the mapped data points with the minimum possible thickness. The mapping to the feature space is obtained implicitly through a kernel function. The formulation is noise resistant and encourages sparsity in the sense only a few "supporting" data points participate in defining the hyperplane, the cost function leads to a quadratic program optimizing a convex function over an union of convex sets. Two alternative formulations are investigated, one where certain $0, 1$ "selection" variables select the constraining convex sets. With the selection variables held fixed the problem reduces to a convex Quadratic Program with closed form updates. The selection variables are optimized by mean field annealing and soft-assign. In the alternative formulation the problem is posed as a Quadratically constrained QP. The optimization problem is reduced to a sequence of one class svm problems using Legendre transformation and solving for the saddle point. Simple generalization error bounds on the probability that a future point drawn iid lands outside of the sandwich is also provided using compression bounds.

given a set of points $\mathbf{x_i} \in \mathbf{R}^n$, we seek projections with low reconstruction error, where the error is defined in a minimax sense. We invoke the margin paradigm of Large Margin methods and define the margin for this *unsupervised* learning task to be the normal distance between two hyperplanes parallel to a subspace, such that all the data points are sandwiched between them. We search for a solution where the margin is the least, or in other words search for a supspace such that the maximum projection distance is minimized.

For an arbitrary lower dimensional subspace given by the matrix W formed by the basis vectors $\mathbf{w_i}$ the cost function can be formulated as:

$$\text{Min}_W \, \text{Max}_i \, ||\mathbf{x_i} - \hat{\mathbf{x}}_\mathbf{i}||^2 \;\; s.t. \;\; \hat{\mathbf{x}}_\mathbf{i} = WW^\dagger \mathbf{x_i} \quad \text{and} \quad \sum_i^k ||\mathbf{w_i}|| = 1 \quad \text{or} \quad \left|\left| W^\dagger W \right|\right|_{tr} = 1. \tag{1}$$