Midterm 2: Advanced Topics in Machine Learning

Instructor: René Vidal

May 21, 2010

1. **K-hyperplanes**: Let $\{\boldsymbol{x}_j \in \mathbb{R}^D\}_{j=1}^N$ be a set of points drawn from a collection of n hyperplanes $\{S_i\}_{i=1}^n$, where $S_i = \{\boldsymbol{x} : \boldsymbol{b}_i^\top \boldsymbol{x} = 0\}$. Propose an iterative algorithm similar to k-means that alternates between assigning points to hyperplanes and estimating the normal to each hyperplane $\boldsymbol{b}_i \in \mathbb{R}^D$. More specifically, write down the cost function to be optimized, the constraints among the optimization variables, and find the optimal normal vectors given the segmentation and the optimal segmentation given the normal vectors.

2. **GPCA.** Let $\{(\boldsymbol{x}_j, \boldsymbol{y}_j)\}_{j=1}^N$ be a set of points in general position drawn from the union of the two surfaces

$$S_1 = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{2D} : \boldsymbol{b}^\top \boldsymbol{x} = 0\} \qquad \text{and} \qquad S_2 = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{2D} : \boldsymbol{y}^\top A \boldsymbol{x} = 0\}, \tag{1}$$

where $\boldsymbol{b} \in \mathbb{R}^D$ and $A \in \mathbb{R}^{D\times D}$ for $D \geq 3$. Assume that $\boldsymbol{b} \neq \boldsymbol{0}$ and that A is full rank.

   (a) Find a polynomial $p(\boldsymbol{x}, \boldsymbol{y})$ that is satisfied by every data point in $S_1 \cup S_2$. Write down a linear system for computing the coefficients of the polynomial $p(\boldsymbol{x}, \boldsymbol{y})$. Find the minimum number of points needed to obtain the coefficients uniquely (up to scale)?

   (b) Show that the clustering of the data points can be obtained by looking at the rank of the following matrix

$$H(\boldsymbol{x}, \boldsymbol{y}) = \frac{\partial^2 p(\boldsymbol{x}, \boldsymbol{y})}{\partial \boldsymbol{x} \partial \boldsymbol{y}} \in \mathbb{R}^{D\times D}. \tag{2}$$

   (c) Derive an algorithm for computing $\boldsymbol{b}$ and A from the derivatives of p at the data points.

3. **SVD-based subspace clustering**: Let $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{D\times N}$ be a matrix whose columns are drawn from a union of n independent linear subspaces$\{S_i\}_{i=1}^n$, i.e.,

$$r = \dim(\cup_{i=1}^n S_i) = \sum_{i=1}^n \dim(S_i) = \sum_{i=1}^n d_i. \tag{3}$$

Let $U_i \in \mathbb{R}^{D\times d_i}$ be a basis for subspace S_i, so that the N_i points in subspace S_i can be written as $X_i = U_i Y_i$, where $Y_i \in \mathbb{R}^{d_i \times N_i}$ is the low-dimensional representation of the data points in subspace S_i. Let $\Gamma \in \mathbb{R}^{N\times N}$ be the unknown permutation matrix sorting the data points according to the subspaces they belong to, i.e., $X = [X_1, X_2, \cdots X_n]\Gamma$. Assume that the points within each subspace are in general position.

   (a) Show that the data matrix X can be factorized as

$$X = \begin{bmatrix} U_1, U_2, \cdots, U_n \end{bmatrix} \begin{bmatrix} Y_1 & 0 & \cdots & 0 \\ 0 & Y_2 & & \\ & & \ddots & 0 \\ 0 & \cdots & 0 & Y_n \end{bmatrix} \Gamma. \tag{4}$$

   (b) Let $M_i \in \mathbb{R}^{N_i \times (N_i - d_i)}$ be a matrix whose columns form an orthonormal basis for the null space of X_i, i.e., $X_i M_i = 0$ and $M_i^\top M_i = I$, and consider the matrix

$$M = \Gamma^\top \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & M_n \end{bmatrix} \in \mathbb{R}^{N\times(N-r)}. \tag{5}$$

   Show that the $N - r$ columns of M are orthonormal and lie in the null space of X.

(c) Let $X = \mathcal{U}\Sigma\mathcal{V}^\top$ be the SVD of the data matrix. Since X has rank r, the matrix $\mathcal{V} \in \mathbb{R}^{N\times N}$ may then be subdivided as $\mathcal{V} = [\mathcal{V}_1\ \mathcal{V}_2]$, where $\mathcal{V}_1 \in \mathbb{R}^{N\times r}$ consists of the first r columns of $\mathcal{V}$ and $\mathcal{V}_2 \in \mathbb{R}^{N\times(N-r)}$ consists of the remaining columns. Show that there exists an orthogonal matrix R such that $\mathcal{V}_2 = MR$.

(d) Recall that $\mathcal{V}\mathcal{V}^\top = \mathcal{V}_1\mathcal{V}_1^\top + \mathcal{V}_2\mathcal{V}_2^\top = I_N$ and let $Q = \mathcal{V}_1\mathcal{V}_1^\top \in \mathbb{R}^{N\times N}$. Show that

$$\Gamma Q \Gamma^\top = I - MM^\top \tag{6}$$

is block diagonal. Use this to show that $Q_{jk} = 0$ if points j and k are in different subspaces.

(e) Suggest an algorithm for clustering independent subspaces based on the SVD of the data matrix X.