# Midterm 1: Advanced Topics in Machine Learning

Instructor: René Vidal

March 5, 2010

1. **PCA.** Let $\boldsymbol{x} \in \mathbb{R}^D$ be a random vector. Let $\mu = \mathbb{E}(\boldsymbol{x}) \in \mathbb{R}^D$ and $\Sigma = \mathbb{E}(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^\top \in \mathbb{R}^{D \times D}$. Define the principal components of $\boldsymbol{x}$ as the random variables $y_i = v_i^\top \boldsymbol{x} + a_i \in \mathbb{R}$, $i = 1, \ldots, d \leq D$, where $v_i \in \mathbb{R}^D$ is a unit norm vector, $a_i \in \mathbb{R}$, and $\{y_i\}$ are zero mean, uncorrelated and of maximum variance $\text{Var}(y_1) \geq \text{Var}(y_2) \geq \cdots \text{Var}(y_d)$. Assuming that the eigenvalues of $\Sigma$ are different from each other, show that

   (a) $a_i = -v_i^\top \mu$, $i = 1, \ldots, d$.

   (b) $v_1$ is the eigenvector of $\Sigma$ corresponding to its largest eigenvalue.

   (c) $v_2^\top v_1 = 0$ and $v_2$ is the eigenvector of $\Sigma$ corresponding to its second largest eigenvalue.

   (d) $v_i^\top v_j = 0$ for all $i \neq j$ and $v_i$ is the eigenvector of $\Sigma$ corresponding to its $i$th largest eigenvalue.

2. **RANSAC.** Imagine we have $\alpha N$ samples from a $d$-dimensional subspace in $\mathbb{R}^D$, where $\alpha \in (0, 1)$. However, the samples are contaminated with $(1 - \alpha)N$ samples that are far from the subspace. We want to estimate the subspace from randomly drawn subsets of $d$ samples. What is the minimum number of subsets that one needs to draw so that, with at least probability $p$, all the points in the subset are inliers?

3. **KPCA.** Consider the polynomial kernel in $[-1, 1]^2 \times [-1, 1]^2$ defined as $k(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^\top \boldsymbol{y})^2 = (x_1 y_1 + x_2 y_2)^2$. Define the operator

$$\mathcal{L}(f)(\boldsymbol{x}) = \int k(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) d\boldsymbol{y}. \tag{1}$$

   (a) Show that $k$ is positive semi-definite.

   (b) Show that the eigenfunctions of $\mathcal{L}$ corresponding to nonzero eigenvalues are of the form $\psi(\boldsymbol{x}) = c_1 x_1^2 + c_2 x_1 x_2 + c_3 x_2^2$. Show that there are three such eigenfuctions, where $(c_1, c_2, c_3)$ and $\lambda$ are obtained from

$$\begin{bmatrix} 4/5 & 0 & 4/9 \\ 0 & 8/9 & 0 \\ 4/9 & 0 & 4/5 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \lambda \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}. \tag{2}$$

4. **LLE.** Let $\{\boldsymbol{x}_i \in \mathbb{R}^D\}_{i=1}^N$ be a collection of data points. Let $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_k}$ be the $k$ nearest neighbors of $\boldsymbol{x}_i$ and define $X_i = [\boldsymbol{x}_{i_1} - \boldsymbol{x}_i, \boldsymbol{x}_{i_2} - \boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i_k} - \boldsymbol{x}_i] \in \mathbb{R}^{D \times k}$. Consider the following optimization problem

$$\min \sum_{i=1}^N \|\boldsymbol{x}_i - \sum_{j=1}^N w_{ij} \boldsymbol{x}_j\|^2 \text{ s.t. } \sum_{j=1}^N w_{ij} = 1,\ i = 1, \ldots, N \text{ and } w_{ij} = 0 \text{ if } \boldsymbol{x}_j \text{ is not a } k \text{ nearest neighbor of } \boldsymbol{x}_i$$

   (a) Show that the cost function can be written as $\sum_{i=1}^N \| \sum_{m=1}^k w_{i,i_m}(\boldsymbol{x}_i - \boldsymbol{x}_{i,m})\|^2 = \sum_{i=1}^N \boldsymbol{w}_i^\top G_i \boldsymbol{w}_i$. where $G_i = X_i^\top X_i \in \mathbb{R}^{k \times k}$ is the Grammian matrix associated with point $i$.

   (b) Show that if $X_i$ is full column rank, the optimal solution to the optimization problem is given by

$$\boldsymbol{w}_i = \begin{bmatrix} w_{i,i_1} \\ w_{i,i_2} \\ \vdots \\ w_{i,i_k} \end{bmatrix} = \frac{G_i^{-1} \mathbf{1}}{\mathbf{1}^\top G_i^{-1} \mathbf{1}} \tag{3}$$

   where $\mathbf{1} \in \mathbb{R}^k$ is the vector of all ones.