Math189R SU17
Homework 5
Monday, June 5, 2017

Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though.

**1 (Murphy 12.5 - Deriving the Residual Error for PCA)** It may be helpful to reference section 12.2.2 of Murphy.

(a) Prove that

$$\left\| \mathbf{x}_i - \sum_{j=1}^{k} z_{ij}\mathbf{v}_j \right\|^2 = \mathbf{x}_i^\top \mathbf{x}_i - \sum_{j=1}^{k} \mathbf{v}_j^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{v}_j.$$

Hint: first consider the case when $k = 2$. Use the fact that $\mathbf{v}_i^\top \mathbf{v}_j$ is 1 if $i = j$ and 0 otherwise. Recall that $z_{ij} = \mathbf{x}_i^\top \mathbf{v}_j$. $z_{ij}^T z_{ij} = v_j^T x_i x_i^T v_j$

(b) Now show that

$$J_k = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^\top \mathbf{x}_i - \sum_{j=1}^{k} \mathbf{v}_j^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{v}_j\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{x}_i - \sum_{j=1}^{k} \lambda_j.$$ eigenvalue of $\Sigma$.

Hint: recall that $\mathbf{v}_j^\top \Sigma \mathbf{v}_j = \lambda_j \mathbf{v}_j^\top \mathbf{v}_j = \lambda_j$.

(c) If $k = d$ there is no truncation, so $J_d = 0$. Use this to show that the error from only using $k < d$ terms is given by

$$J_k = \sum_{j=k+1}^{d} \lambda_j.$$

Hint: partition the sum $\sum_{j=1}^{d} \lambda_j$ into $\sum_{j=1}^{k} \lambda_j$ and $\sum_{j=k+1}^{d} \lambda_j$.

a. $\|\vec{x}_i - \sum_{j=1}^{k} z_{ij}\vec{v}_j\|^2 = (\vec{x}_i - \sum_{j=1}^{k} z_{ij}\vec{v}_j)^T(\vec{x}_i - \sum_{j=1}^{k} z_{ij}\vec{v}_j)$

$= \vec{x}_i^T\vec{x}_i - 2\sum_{j=1}^{k} z_{ij}\vec{v}_j^T\vec{x}_i + (\sum_{j=1}^{k} z_{ij}\vec{v}_j)^T(\sum_{j=1}^{k} z_{ij}\vec{v}_j)$

$= \vec{x}_i^T\vec{x}_i - 2\sum_{j=1}^{k} z_{ij}\vec{v}_j^T\vec{x}_i + \sum_{j=1}^{k} \vec{v}_j^T z_{ij}^T z_{ij}\vec{v}_j$

$= \vec{x}_i^T\vec{x}_i - 2\sum_{j=1}^{k} \vec{v}_j^T\vec{x}_i\vec{x}_i^T\vec{v}_j + \sum_{j=1}^{k} v_j^T x_i x_i^T v_j$

$= x_i^T x_i - \sum v_j^T x_i x_i^T v_j$

■

b. $J_k = \frac{1}{n}\sum_{i=1}^{n} \left(x_i^T x_i - \sum_{j=1}^{k} v_j^T x_i x_i^T v_j\right)$

$= \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} v_j^T \frac{1}{n}\left(\sum_{i=1}^{n} x_i x_i^T\right) v_j$

$= \cdots\cdots - \sum_{j=1}^{k} v_j^T \Sigma v_j$

$= \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} \lambda_j$

c. $\sum_{j=1}^{d} \lambda_j = \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i$

So $J_k = \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{d} \lambda_j + \sum_{j=k+1}^{d} \lambda_j = \sum_{j=k+1}^{d} \lambda_j$.

**2 ($\ell_1$-Regularization)** Consider the $\ell_1$ norm of a vector $\mathbf{x} \in \mathbb{R}^n$:

$$\|\mathbf{x}\|_1 = \sum_i |\mathbf{x}_i|.$$

Draw the norm-ball $B_k = \{\mathbf{x} : \|\mathbf{x}\|_1 \leq k\}$ for $k = 1$. On the same graph, draw the Euclidean norm-ball $A_k = \{\mathbf{x} : \|\mathbf{x}\|_2 \leq k\}$ for $k = 1$ behind the first plot. (Do not need to write any code, draw the graph by hand).

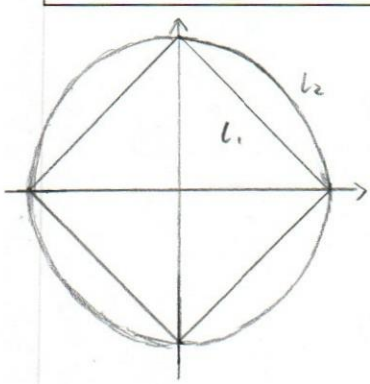
Show that the optimization problem

$$\begin{aligned} &\text{minimize: } f(\mathbf{x}) \\ &\quad \text{subj. to: } \|\mathbf{x}\|_p \leq k \end{aligned}$$

is equivalent to

$$\text{minimize: } f(\mathbf{x}) + \lambda\|\mathbf{x}\|_p$$

(hint: create the Lagrangian). With this knowledge, and the plots given above, argue why using $\ell_1$ regularization (adding a $\lambda\|\mathbf{x}\|_1$ term to the objective) will give sparser solutions than using $\ell_2$ regularization for suitably large $\lambda$.




The question is equivalent to

$$\inf_x \sup_{\lambda \geq 0} L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} f(x) + \lambda(\|x\|_p - k).$$

$$\Rightarrow \sup_{\lambda \geq 0} \inf_x f(x) + \lambda(\|x\|_p - k) = \sup_{\lambda \geq 0} g(\lambda).$$

The min of $f(x) + \lambda(\|x\|_p - k)$ over $x$ is the same as min of $f(x) + \lambda\|x\|_p$.

Since $l_1$ has sharp corners, the possibility of the optimum point landing on the corner is infinitely larger than that of landing on the edge.

∎