Cassidy Lê
Math189R SP19
Homework 5
Monday, Mar 4, 2019

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

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

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

(b) Now show that

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

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

(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) With the help of section 12.2.2 of Murphy, we know that

$$\left\| \mathbf{x}_i - \sum_{j=1}^{k} z_{ij}\mathbf{v}_j \right\|^2 = \left(x_i - \sum_{j=1}^{k} z_{ij}\mathbf{v}_j\right)^T\left(x_i - \sum_{j=1}^{k} z_{ij}\mathbf{v}_j\right)$$

$$= x_i^T x_i - x_i^T\sum_{j=1}^{k} z_{ij}\mathbf{v}_j - x_i\sum_{j=1}^{k} z_{ij}\mathbf{v}_j^T + \left(\sum_{j=1}^{k} z_{ij}\mathbf{v}_j\right)^T\left(\sum_{j=1}^{k} z_{ij}\mathbf{v}_j\right).$$

Rearranging using transpose properties,

$$= x_i^T x_i - 2\sum_{j=1}^{k} z_{ij}\mathbf{v}_j^T x_i + \sum_{j=1}^{k}\mathbf{v}_j^T z_{ij}^T z_{ij}\mathbf{v}_j.$$

Since $\mathbf{v}_i^T \mathbf{v}_j$ is 1 if $i = j$ and 0 otherwise and $z_{ij} = x_I^T \mathbf{v}_j$,

$$= x_i^T x_i - 2 \sum_{j=1}^{k} z_{ij} \mathbf{v}_j^T x_i + \sum_{j=1}^{k} \mathbf{v}_j^T x_i x_i^T \mathbf{v}_j$$

$$= x_i^T x_i - 2 \sum_{j=1}^{k} \mathbf{v}_j^T x_i x_i^T \mathbf{v}_j + \sum_{j=1}^{k} \mathbf{v}_j^T x_i x_i^T \mathbf{v}_j$$

$$\left\| \mathbf{x}_i - \sum_{j=1}^{k} z_{ij} \mathbf{v}_j \right\|^2 = x_i^T x_i - \sum_{j=1}^{k} \mathbf{v}_j^T x_i x_i^T \mathbf{v}_{j \cdot QED}$$

(b) By definition, it is true that

$$J_k = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{x}_i - \sum_{j=1}^{k} \mathbf{v}_j^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{v}_j \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{k} \mathbf{v}_j^T \frac{1}{n} \left(\sum_{j=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{v}_j$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{k} \mathbf{v}_j^T \Sigma \mathbf{v}_j$$

Since $\mathbf{v}_j^T \Sigma \mathbf{v}_j = \lambda_j$,

$$J_k = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{k} \lambda_{j \cdot QED}$$

(c) And from part (b), we know that $J_k = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{k} \lambda_j$. Additionally, we can partition $\sum_{j=1}^{d} \lambda_j$ such that

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

If we rearrange this equation, we get

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

We can substitute this into the equation of J_k, so

$$J_k = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{d} \lambda_j + \sum_{i=k+1}^{d} \lambda_j$$

If $J_d = 0$, then it must be true that

$$\sum_{k=1}^{d} \lambda_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i.$$

Therefore,

$$J_k = \sum_{i=k+1}^{d} \lambda_{j \cdot QED}$$

■

2 (ℓ_1-Regularization) Consider the ℓ_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 ℓ_1 regularization (adding a $\lambda\|\mathbf{x}\|_1$ term to the objective) will give sparser solutions than using ℓ_2 regularization for suitably large λ.

The graph can be seen in **Figure 1** on the next page. The optimization problem of minimizing $f(\mathbf{x})$ subject to $||x||_p \leq k$ is the same as the following optimization problem:

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

By switching the supremums and infemums, we can an equivalent of

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

Because the term λk does not depend on $\mathbf{x}$, minimizing all of $f(\mathbf{x}) + \lambda(||\mathbf{x}||_p - k)$ is equivalent to minimizing $f(\mathbf{x}) + \lambda||\mathbf{x}||_p$. Therefore, the two optimization problems are equivalent.$_{QED}$ ∎

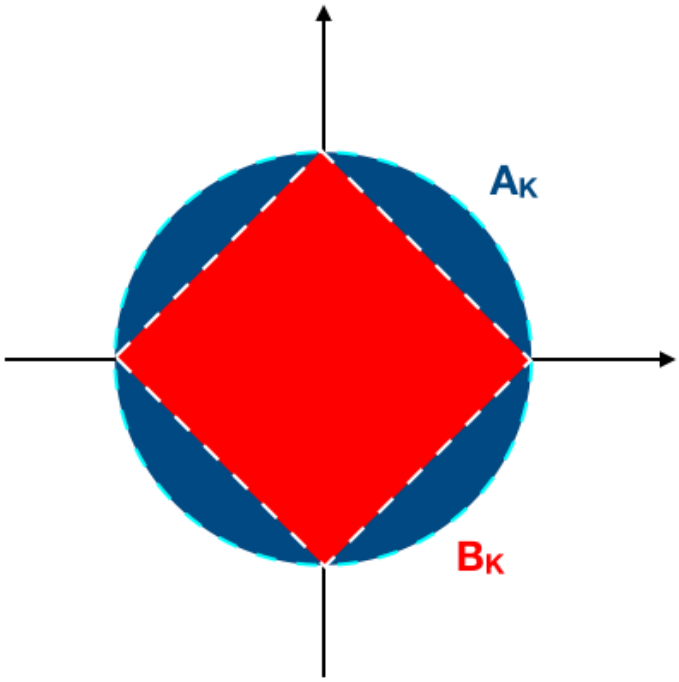

Figure 1: ℓ_1 Norm-Ball and Euclidean Norm-Ball