CS4786/5786: Machine Learning for Data Science, Spring 2015
2/19/2015: Updated "cheatsheet" on projections

1 Recap

We will use bold letters to denote vectors. For instance, $\mathbf{x} \in \mathbb{R}^d$ means $\mathbf{x}$ is a d-dimensional vector whose coordinates are real-valued. This vector can be represented as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}[1] \\ \mathbf{x}[2] \\ \vdots \\ \mathbf{x}[d] \end{bmatrix}$$

where $\mathbf{x}[i]$ is the value on the ith coordinate. We shall use capital letters to denote matrices. For instance, if W is a $d \times k$ matrix then $W[i, j]$ represents the entry (i, j). We can write the matrix W as $W = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$ where each $\mathbf{w}_i$ is a d-dimensional vector.

Recall the notion of Euclidean norm $\|\cdot\|_2$. Notice that for any vector $\mathbf{x} \in \mathbb{R}^d$,

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{d} \mathbf{x}[i]^2 = \mathbf{x}^\top \mathbf{x}.$$

The norm of a vector is the distance of this vector from the origin.

Given a vector $\mathbf{x}$, the outer product $\mathbf{x}\mathbf{x}^\top$ yields a $d \times d$ matrix with entries being all the products of pairs of entries of $\mathbf{x}$:

$$\mathbf{x}\mathbf{x}^\top = \begin{bmatrix} \mathbf{x}[1] \cdot \mathbf{x}[1], & \mathbf{x}[1] \cdot \mathbf{x}[2], & \ldots, & \mathbf{x}[1] \cdot \mathbf{x}[d] \\ \mathbf{x}[2] \cdot \mathbf{x}[1], & \mathbf{x}[2] \cdot \mathbf{x}[2], & \ldots, & \mathbf{x}[2] \cdot \mathbf{x}[d] \\ & & \vdots & \\ \mathbf{x}[d] \cdot \mathbf{x}[1], & \mathbf{x}[d] \cdot \mathbf{x}[2], & \ldots, & \mathbf{x}[d] \cdot \mathbf{x}[d] \end{bmatrix}$$

The data matrix X is an $n \times d$ matrix given by

$$X = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$$

Henceforth, we shall assume that the $\mathbf{x}$*'s are centered, that is,* $\frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t = \mathbf{0}$.

For centered $\mathbf{x}$'s, the covariance matrix is given by

$$\Sigma = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t^\top = \frac{1}{n} X^\top X$$

[original text had $X^\top$ after X]

PCA : Find top K eigen vectors of Σ, given by $\mathbf{w}_1, \ldots, \mathbf{w}_K \in \mathbb{R}^d$. Set

$$W = \begin{bmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_K^\top \end{bmatrix}.$$

[The previous cheatsheet handed out used the transpose of this. Doing so can be mathematically fine (if you're consistent all the way through)! But here we'll switch back to the conventions set up in lecture 2, where the eigenvectors we're interested in are considered to make up the rows of W.] Now the way we find the low dimensional representation is as

$$\mathbf{y}_t = W\mathbf{x} \qquad (K \times d \text{ and } d \times 1 \Rightarrow K \times 1)$$

[original text: $\mathbf{y}_t^\top = \mathbf{x}^\top W$] or $Y = WX^\top$ ($K \times d$ and $d \times n \Rightarrow K \times n$) [original: $Y = X \times W$].

CCA : Given two views of same n data points, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{x}'_1, \ldots, \mathbf{x}'_n$ where the $\mathbf{x}_t$'s are d-dimensional and the $\mathbf{x}'_t$'s are d'-dimensional, find corresponding low-dimensional representations $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^K$ and $\mathbf{y}'_1, \ldots, \mathbf{y}'_n \in \mathbb{R}^K$ based on common information between both the views. Here, we are considering the columns of W and V below to represent the vectors to be projected on to.

$$\mathbf{y}_t^\top = \mathbf{x}_t^\top W \quad \text{and} \quad \mathbf{y'}_t^\top = \mathbf{x'}_t^\top V$$

are linear transformations of the two views where W and V are jointly calculated based on both views.

Procedure for CCA:

1. Make the concatenated view
$$\tilde{\mathbf{x}}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}'_t \end{bmatrix}$$
2. Calculate covariance matrix of $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$, call it $\tilde{\Sigma}$. This is a $(d + d') \times (d + d')$ matrix
3. Pick out $\Sigma_{X,X}$ the covariance for view 1 which is a $d \times d$ matrix, $\Sigma_{X',X'}$ the covariance for view 2 which is a $d' \times d'$ matrix, $\Sigma_{X,X'}$ the covariance of view 1 w.r.t. view 2 which is a $d \times d'$ matrix, as
$$\tilde{\Sigma} = \begin{bmatrix} \Sigma_{X,X} & \Sigma_{X,X'} \\ \Sigma_{X',X} & \Sigma_{X',X'} \end{bmatrix}$$
4. The columns of W are the top K eigenvectors of the $d \times d$ matrix $\Sigma_{X,X}^{-1}\Sigma_{X,X'}\Sigma_{X',X'}^{-1}\Sigma_{X',X}$. Similarly, the columns of V are the top K eigenvectors of $d' \times d'$ matrix $\Sigma_{X',X'}^{-1}\Sigma_{X',X}\Sigma_{X,X}^{-1}\Sigma_{X,X'}$.

2 Random Projections

Fill the $d \times K$ matrix W as follows. Here, we're considering the columns as the vectors to project on to. For each entry (i, j), flip a fair coin; if heads, enter $+1/\sqrt{K}$, whereas if tails, enter $-1/\sqrt{K}$. In the Matlab code demoed in class, the "sign" function should be applied to the output of `randn` to make sure one gets either plus-one or minus-one as a result for each entry's numerator. Notice two things.

- Since each entry is independently sampled, given the generation process outlined above, for any $i \neq i'$ and any j,

$$\mathbb{E}\left[W[i, j] \cdot W[i', j]\right] = 0$$

- Further

$$W[i, j] \cdot W[i, j] = W[i, j]^2 = \left(\frac{\pm 1}{\sqrt{K}}\right)^2 = \frac{1}{K}$$