P-Set #1

David Sideman AMATH 797 01/27/20

(1) We want to project $X \in \mathbb{R}^{p \times n}$ onto a lower dimensional subspace $V \in \mathbb{R}^{p \times d}$. The optimal projection is the orthogonal projection (Gram-Schmidt) given by $V^T X$. Therefore, this implies $\boxed{\beta_k = V^T x_k}$

$$\min_{\substack{\mu, V, \beta \\ V^TV = I}} \sum_{k=1}^{n} \|x_k - \mu + V\beta_k\|_2^2 = \sum_{k=1}^{n} \|x_k - \mu + V(V^T x_k)\|_2^2$$

$$= \sum_{k=1}^{n} \|x_k\|^2 + \|\mu + VV^T x_k\|^2 - 2\langle x_k, \mu + VV^T x_k\rangle$$

$$= \sum_{k=1}^{n} \|x_k\|^2 + \|\mu + VV^T x_k\|^2 - 2\langle x_k, \mu\rangle - 2\langle x_k, VV^T x_k\rangle$$

$$= \sum_{k=1}^{n} \|x_k\|^2 + \|\mu\|^2 + \|VV^T x_k\|^2 + 2\langle \mu, VV^T x_k\rangle - 2\langle x_k, \mu\rangle - 2\langle x_k, VV^T x_k\rangle$$

Take partial to optimize

$$= \sum_{k=1}^{n} \frac{\partial}{\partial \mu}\left(\|\mu\|^2 + 2\langle \mu, VV^T x_k\rangle - 2\langle x_k, \mu\rangle + \|x_k\|^2 + \|VV^T x_k\|^2 - 2\langle x_k, VV^T x_k\rangle\right) = 0$$

$$= \sum_{k=1}^{n} 2\mu + 2VV^T x_k - 2x_k = 0$$

$$= \sum_{k=1}^{n} \mu = \sum_{k=1}^{n} (I - VV^T) x_k$$

$$= n\mu = (I - VV^T) \sum_{k=1}^{n} x_k$$

$$\Rightarrow \boxed{\mu = \frac{1}{n} \sum_{k=1}^{n} x_k}$$

; **Lemma**: $(I - VV^T)(I - VV^T) = I - 2VV^T + VV^TVV^T = I - VV^T$ $\therefore$ Idempotent $\Rightarrow$ psd $\Rightarrow$ convex

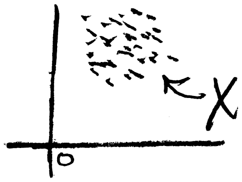
Thus choosing μ to be the sample mean is a valid solution, and due to convexity, all solutions are equivalent minima

$$\min_{\substack{\mu, V, \beta \\ V^TV = I}} \sum_{k=1}^{n} \|x_k - \left(\tfrac{1}{n}\sum_{k=1}^{n} x_k\right) + VV^T x_k\|_2^2 = \sum_{k=1}^{n} \|(x_k - \mu_n) + VV^T x_k\|_2^2$$

$$= \sum_{k=1}^{n} \|x_k - \mu_n\|^2 + \|VV^T x_k\|^2 - 2\langle x_k - \mu_n, VV^T x_k\rangle = \|x_k - \mu_n\| + (VV^T x_k)^T (VV^T x_k) - 2(x_k - \mu_n)^T (VV^T x_k)$$

$$= \sum_{k=1}^{n} \|x_k - \mu_n\|^2 - (x_k - \mu_n)^T VV^T (x_k - \mu_n) = \|x_k - \mu_n\|^2 - \mathrm{trace}\left(V^T (x_k - \mu_n)(x_k - \mu_n)^T V\right)$$

Minimize Quantity: Constant Maximizing this trace is equivalent to PCA.

② 

Clearly, the mean of the data scattered in the first quadrant is not the origin. Therefore, without being centered at the origin and having expectation 0, the covariance matrix of X:

$$(X-\mathbb{E}X)(X-\mathbb{E}X)^T = XX^T \quad \text{iff } \mathbb{E}X = 0$$

Thus, finding the eigenvectors of XX^T for the left U matrix of SVD is <u>NOT</u> equivalent to finding the eigenvectors of X's covariance matrix as is done in PCA.

③ <u>Mean</u>: $\frac{1}{n}\sum_{k}^{n}\beta_i = \frac{1}{n}\sum_{k}^{n} U_i^T(X-\mu_s 1^T) = U_i^T \frac{1}{n}\sum_{k}^{n}\begin{bmatrix} | & & | \\ (x_1-\mu_s) & \cdots & (x_n-\mu_s) \\ | & & | \end{bmatrix}$

$= U_i^T \frac{1}{n}\begin{bmatrix} \sum_k^n (x_{1k}-\mu_s) \\ \vdots \\ \sum_k^n (x_{dk}-\mu_s) \end{bmatrix}$, By definition of sample mean $= U_i^T \frac{1}{n}\begin{bmatrix} - 0_1 - \\ \vdots \\ - 0_d - \end{bmatrix} = 0$

<u>Uncorrelated</u>: $\beta\beta^T = \left(U^T(X-\mu_s 1^T)\right)\left(U^T(X-\mu_s 1^T)\right)^T = U^T(X-\mu_s 1^T)(X-\mu_s 1^T)^T U$

$= U^T \Sigma_s U$

$= \frac{S^2}{n-1}$

(I'm assuming U are principal directions, not principal components, or dimensions break)

and we know that the principal directions which are the eigenvectors of the covariance matrix diagonalize Σ.
$\therefore$ the off-diagonals of the covariance matrix for β are all 0, so any pair of meta-features are uncorrelated.