1. Residual error for PCA.

Oscar Scholin

189 HW 5

a) Prove $\left\| \vec{x}_i - \sum_{j=1}^{K} z_{ij} \vec{v}_j \right\|^2 = \vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \vec{v}_j^T \vec{x}_i \vec{x}_i^T \vec{v}_j$.

Apply definition of the norm:

$\langle A, B \rangle = \overline{\langle B, A \rangle}$

$= \langle B, A \rangle$ if real vector space

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

$$= \vec{x}_i^T \vec{x}_i \underbrace{- \sum_{j=1}^{K} z_{ij}^T \vec{v}_j^T \vec{x}_i - \vec{x}_i^T \sum_{j=1}^{K} z_{ij}^T \vec{v}_j} + \left(\sum_{j=1}^{K} z_{ij} \vec{v}_j\right)^T \sum_{j=1}^{K} z_{ij} \vec{v}_j$$

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

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

since $z_{ij} \in \mathbb{R}$

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

$= z_{ij} z_{ij}^T$

$= \vec{v}_j^T \vec{x}_i^T \vec{v}_j \vec{x}_i \vec{v}_j^T \vec{v}_j$

$= \vec{v}_j^T \vec{v}_j \vec{x}_i^T \vec{x}_i \vec{v}_j^T \vec{v}_j$

$= \vec{v}_j^T \vec{v}_j \vec{x}_i \vec{x}_i^T \vec{v}_j^T \vec{v}_j$ since $\vec{v}_j^T \vec{v}_j = 1$

$= \vec{x}_i \vec{x}_i^T$

$= v_j^T v_j \vec{x}_i \vec{x}_i^T$

$= v_j^T \underbrace{\vec{x}_i \vec{x}_i^T}_{\text{since } \in \mathbb{R}} v_j$

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

$$= \vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \vec{v}_j^T \vec{x}_i \vec{x}_i^T \vec{v}_j \quad \square$$

reconstruction using 1st k principal components

b) Show $J_K = \frac{1}{n} \sum_{i=1}^{n} \left(\vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \vec{v}_j^T \vec{x}_i \vec{x}_i^T \vec{v}_j\right) = \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \lambda_j$

Recall $\vec{v}_j^T \Sigma \vec{v}_j = \lambda_i \vec{v}_j^T \vec{v}_j = \lambda_j$

Reconstruction error

By definition,

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

$$= \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \vec{v}_j^T \underbrace{\frac{1}{n}\left(\sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T\right)}_{\equiv \Sigma} \vec{v}_j$$

$$= \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \vec{v}_j^T \Sigma \vec{v}_j$$

using the definition

$$= \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i^T \vec{x}_i - \sum_{j=1}^{K} \lambda_j$$

c) If $k = d$, there is no truncation.

$$\therefore \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i^{\,T} \vec{x}_i = \sum_{j=1}^{d} \lambda_j$$

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

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

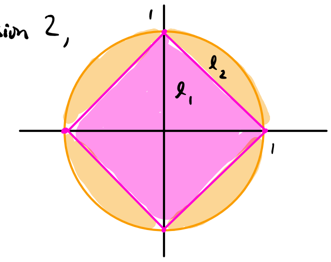
2. l_1-regularization. Consider l_1-norm of $\vec{x} \in \mathbb{R}^n$:

$\|\vec{x}\|_1 = \sum_i |\vec{x}_i|$

Draw the norm ball $B_k = \{\vec{x} : \|\vec{x}\|_1 \le k\}$ for $k = 1$ norm ball

$A_k = \{\vec{x} : \|\vec{x}\|_2 \le k\}$ Euclidean norm ball

for dimension 2,



$B_k : \|\vec{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$ $\Rightarrow$ forms diamond since moving horizontally or vertically from origin increases norm by same as moving directly away

$A_k : \|\vec{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ $\Rightarrow$ forms circle since distance from origin to any point on circle

Show optimization problem: minimize $f(\vec{x})$ subject to $\|\vec{x}\|_p \le k$ is equivalent to $f(\vec{x}) + \lambda\|\vec{x}\|_p$

We know minimize $f(\vec{x})$ subject to $\|\vec{x}\|_p \le k$ is equivalent to

Lagrangian. This is the primal problem.

$$\inf_{\vec{x}} \sup_{\lambda \ge 0} \mathcal{L}(\vec{x}, \lambda) = \inf_{\vec{x}} \sup_{\lambda \ge 0} \underbrace{f(\vec{x}) + \lambda \overbrace{(\|\vec{x}\|_p - k)}^{\text{constraint}}}_{\mathcal{L}(\vec{x}, \lambda)}$$

Note we can flip the infimum and supremum and make it a dual problem:

$$\sup_{\lambda \ge 0} \inf_{\vec{x}} f(\vec{x}) + \lambda(\|\vec{x}\|_p - k) = \sup_{\lambda \ge 0} g(\lambda)$$

Note minimizing $f(\vec{x}) + \lambda(\|\vec{x}\|_p - k)$ over $\vec{x}$ is equivalent to minimizing $f(\vec{x}) + \lambda\|\vec{x}\|_p$ $\square$

If we use l_1 norm, there is a greater likelihood that the optimized solution lies on an axis since the norm ball intersects the axes with sharp corners.
However, with an l_2 norm, the solution is less likely to lie on the axis. Therefore we are more likely to find solutions with some components exactly 0 whereas w/ an l_2 norm they are more likely to be small but non-0.