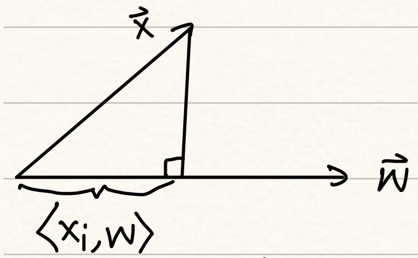
Lec 5: PCA and SVD

$\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n \in \mathbb{R}^n$ "lower dim structure"

Goal: Find $\vec{w} \in \mathbb{R}^p$ s.t. the projections of our data onto $\vec{w}$ are as close to the original vectors as possible, $\|\vec{w}\|_2 = 1$, zero-mean $\left(\frac{1}{n}\sum \vec{x}_i = 0\right)$

Projection of $\vec{x}_i$ onto $\vec{w}$



$$(A^TA)^{-1}A^T\vec{b} = (\vec{w}^T\vec{w})^{-1}\vec{w}^T\vec{x}_i$$

$$e_i^2 = \|\vec{x}_i - \langle \vec{x}_1, \vec{w}\rangle \vec{w}\|^2$$

$$\underset{\vec{w}}{\text{Minimize}} \ \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Mean of projections

$$\frac{1}{n}\sum_{i=1}^{n}\langle \vec{x}_i, \vec{w}\rangle \vec{w}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i^T\vec{w})\vec{w} = \frac{1}{n}\sum(\vec{w}^T\vec{x}_i)\vec{w}$$

$$= \Big(\big(\underbrace{\tfrac{1}{n}\sum \vec{x}_i^T}_{0}\big)\vec{w}\Big)\vec{w}$$

$$= 0$$

$$e_i^2 = \|\vec{v}_i\|^2 = \vec{\theta}_i^T\vec{\theta}_1$$

$$= (\vec{x} - \langle \vec{x}_i, \vec{w}\rangle \vec{w})^T(\vec{x}_i - \langle \vec{x}_i, \vec{w}\rangle \vec{w})$$

$$= \|\vec{x}_i\|_2^2 - 2\vec{x}_i^T w\langle \vec{x}_i, \vec{w}\rangle + \langle \vec{x}_i, w\rangle^2 \|\vec{w}\|_2^2$$

$$= \|\vec{x}_i\|_2^2 - 2\langle \vec{x}_i, \vec{w}\rangle^2 + \langle \vec{x}_i, \vec{w}\rangle^2$$

$$= \|\vec{x}_i\|_2^2 - \langle \vec{x}_i, \vec{w}\rangle^2$$

$$\min_{\substack{\vec{w} \\ \|\vec{w}\|_2^2=1}} \frac{1}{n}\sum_{i=1}^{n}\left(-\langle x_i, \vec{w}\rangle^2\right) \Rightarrow \max_{\substack{\vec{w} \\ \|\vec{w}\|_2^2=1}} \frac{1}{n}\sum_{i=1}^{n}\langle x_i, \vec{w}\rangle^2$$ (removed negative, flipped to max)

$$\frac{1}{n}\sum_{i=1}^{n}\langle \vec{x}_i, \vec{w}\rangle^2 = \underbrace{\left(\frac{1}{n}\sum\langle \vec{x}_i, \vec{w}\rangle\right)^2}_{0} + \text{Var}$$

doesn't need to be $\frac{1}{n^2}$ since it is a constant, maximizing value will be the same

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\langle \vec{x}_i, \vec{w}\rangle^2 &= \frac{1}{n}\sum \vec{w}^T\vec{x}_i\,\vec{x}_i^T\vec{w} \\
&= \frac{1}{n} w^T\left(\sum \vec{x}_i\,\vec{x}_i^T\right)\vec{w} \\
&= \frac{1}{n}\vec{w}^T(X^TX)\vec{w} \\
&= \vec{w}\left(\frac{X^TX}{n}\right)\vec{w}
\end{aligned}$$

Rayleigh coeff

covariance matrix symmetric

$$\max_{\|\vec{w}\|_2=1} \vec{w}^TA\vec{w}, \quad A \in S^p$$

By the spectral theorem, $A = U^T\Lambda U$, U is orthogonal, Λ diagonal

$$= \max_{\|\vec{w}\|_2=1} w^T U^T \Lambda U^T \vec{w}$$

$$\|U\vec{x}\|_2^2 = \|x^TU^TU\vec{x}\|_2 = \|\vec{x}\|_2^2$$

$$= \max_{\|y\|_2^2=1} \vec{y}^T\Lambda\vec{y}, \quad y = U\vec{w}$$

$$\lambda_{min} \le \sum \lambda_i y_i^2 \le \lambda_{max}$$

$$= \max_{\|y\|_2^2=1} \sum \lambda_i y_i^2$$

$$\lambda_{min} \le \underbrace{\frac{\vec{X}^TA\vec{X}}{\|X\|_2^2}}_{\text{rayleigh coeff}} \le \lambda_{max}$$

$$\min \frac{1}{n}\sum e_i^2 \Leftrightarrow \max_{\|\vec{w}\|=1} \sum\langle \vec{x}_i, \vec{w}\rangle^2 \Leftrightarrow \max_{\|w\|_2^2=1} \vec{w}^TA\vec{w} = \lambda_{max}(A)$$

argmax $\rightarrow \vec{w}^* = \text{eigenvector}(\lambda_{max})$

SVD

$A \in \mathbb{R}^{m,n}$

$A = U\Sigma V^T$, U, V orthonormal, Σ almost diagonal

$$\boxed{A} = \boxed{U} \begin{bmatrix} \Sigma_r \\ 0 \end{bmatrix} \boxed{V^T}$$

"Full SVD"

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^{\,T}, \quad A \text{ rank } r$$

Compact SVD: $A = U_r \Sigma_r V_r^T$