Agenda:
- HW13 due 4/29
- Course Evals - EC opportunity
- PCA
- Worksheet

Low-rank Approximation

Given matrix $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = r$

→ Want: $A_\ell \in \mathbb{R}^{m \times n}$, $\text{rank}(A_\ell) = \ell << r$ that is "close" to A ($\|A - A_\ell\|$ is small)

Recall Outer Product Form of SVD:

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T \quad (\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots)$$

$$= \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^T + \underbrace{\sum_{i=\ell+1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T}_{\sigma_i \approx 0}$$

$$\approx \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^T$$

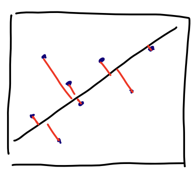
$$A_\ell = \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^T$$

Principal Component Analysis: captures most important dimensions of data

Let $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^d$ be data points. If $A = [\vec{x}_1, \dots \vec{x}_n]$ w/ SVD $A = U \Sigma V^T$, then

$$U_\ell \in \underset{W \in \mathbb{R}^{d \times \ell}}{\text{argmin}} \sum_{i=1}^{n} \|\vec{x}_i - WW^T \vec{x}_i\|^2$$

where $U_\ell = [\vec{u}_1 \dots \vec{u}_\ell]$ is the 1st ℓ columns of U



→ Geometric interpretation:

$$S_{PCA} \in \underset{S \subseteq \mathbb{R}^d}{\text{argmin}} \sum_{i=1}^{n} \|\vec{x}_i - \text{proj}_S(\vec{x}_i)\|^2 \quad \text{s.t. } \dim(S) \leq \ell$$

$\vec{x}_i - \text{proj}_S(\vec{x}_i)$

↳ trying to minimize this error across all data points

Meaning: U_ℓ is a set of ℓ orthonormal vectors in $\mathbb{R}^m$ that finds the "best" subspace that is closest to all points

Def: Principal Components

$A = [\vec{x}_1, \dots, \vec{x}_n]$, data stored in columns

⇒ PC of A are eigenvectors of AA^T, ordered in non-decreasing order by value of corresponding eigenvalue

⇒ U_ℓ stores principal components

★ If data stored in rows, then

⇒ PC of A are eigenvectors of A^TA, ordered by value of corresponding eigenvalue

⇒ V_ℓ stores principal components

Intuition: Back to outer products

$$\sigma_i \vec{u}_i \vec{v}_i^T = \sigma_i \begin{bmatrix} | \\ \vec{u}_i \\ | \end{bmatrix} [- \vec{v}_i^T -]$$

2 interpretations:

(1) $\vec{u}_i$ is the data and you are scaling each column by the components of $\vec{v}_i^T$

⇒ interpreting data in columns

(2) $\vec{v}_i^T$ is the data & you are scaling each row by the components of $\vec{u}_i$

⇒ interpreting data in rows

→ σ_i dictates how much this outer product $\vec{u}_i \vec{v}_i^T$ contributes to overall data matrix A