Announcements

- Milestone due tonight
- Fill in the missing plots:

\[ X = \sum_{k=1}^{n} u_k v_k^T s_k \]

\[ JX = X - \frac{1}{n} X^T \]

\[ U, S, V = \text{svd}(X) \Rightarrow U^T U = I, V^T V = I \]

\[ JX = U S V^T \]

\[ J = I - \frac{11^T}{n} \]

\[ JXV^S^{-1} \]

\[ JXV^S^{-1} V^T \]

\[ JXV^S^{-1} = USV^TV^S^{-1} = U \]
Linear projections

Given $x_i \in \mathbb{R}^d$ and some $q < d$ consider

$$\min_{V_q} \sum_{i=1}^{N} \|(x_i - \bar{x}) - V_q V_q^T (x_i - \bar{x})\|^2.$$ 

where $V_q = [v_1, v_2, \ldots, v_q]$ is orthonormal: 

$$V_q^T V_q = I_q$$

$V_q$ are the first $q$ eigenvectors of $\Sigma$

$V_q$ are the first $q$ principal components

Principal Component Analysis (PCA) projects $(X - 1\bar{x}^T)$ down onto $V_q$

$$(X - 1\bar{x}^T)V_q = U_q \text{diag}(d_1, \ldots, d_q)$$

$$U_q^T U_q = I_q$$

Singular Value Decomposition defined as

$$X - 1\bar{x}^T = U S V^T$$
Linear projections

$X$ is centered ($TX = x$)

$XX^T = US^2U^T$

$eig(XX^T) = U$

$XTX = VS^2V^T$

$eig_{vec}(XTX) = V$

$A = XV$

$u_i = \frac{Ae_i}{||Ae_i||_2} = u_i \frac{s_i}{s_i} = u_i$

$Ae_i = u_i s_i$

$||Ae_i||_2^2 = s_i^2 ||u_i||_2^2 = s_i^2$
Dimensionality reduction

\( V_q \) are the first \( q \) eigenvectors of \( \Sigma \) and SVD \( X - 1\bar{x}^T = USV^T \)
Power method - one at a time

\[ \Sigma := \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \quad v_* = \arg \max_v v^T \Sigma v \]
Power method - one at a time

\[ \Sigma := \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \]

\[ v_* = \arg \max_v v^T \Sigma v \]

\[ v_k^{h+1} = \frac{\Sigma v_k}{\| \Sigma v_k \|} \]

\[ \sum_{k=0}^{h} (USU^T)^k = VS^h V^T \]

\[ \frac{US^h V}{\| US^h V \|_2} = \frac{s^h V \begin{bmatrix} (s, s, \ldots)^k \\ \vdots \\ (s, s, \ldots)^k \end{bmatrix} \alpha}{s^h \| V \begin{bmatrix} (s, s, \ldots)^k \\ \vdots \\ (s, s, \ldots)^k \end{bmatrix} \alpha \|_2} \longrightarrow \frac{s^h V \begin{bmatrix} \alpha \end{bmatrix}}{s^h \| V \begin{bmatrix} \alpha \end{bmatrix} \|} = V_i \]
Markov chains - PageRank
Markov chains - PageRank

$L_{i,j} = 1\{\text{page } j \text{ points to page } i\}$

Google PageRank of page $i$:

$$p_i = (1 - \lambda) + \lambda \sum_{j=1}^{n} \frac{L_{i,j}}{c_j} p_j$$

$$c_j = \sum_{k=1}^{n} L_{j,k}$$
Markov chains - PageRank

$L_{i,j} = 1 \{ \text{page } j \text{ points to page } i \}$

Google PageRank of pages given by:

$$p = (1 - \lambda)1 + \lambda LD_c^{-1}p$$

$D_c = \text{diag}(c_1, \ldots, c_n)$
$L_{i,j} = 1$\{page $j$ points to page $i$\}

Google PageRank of pages given by:

$$p = (1 - \lambda)1 + \lambda LD_c^{-1}p$$

Set arbitrary normalization: $1^T p = n$ so that

$$p = ((1 - \lambda)11^T / n + \lambda LD_c^{-1})p$$

$=: Ap$
Markov chains - PageRank

$L_{i,j} = 1\{\text{page } j \text{ points to page } i\}$

Google PageRank of pages given by:

$$p = (1 - \lambda)1 + \lambda LD_c^{-1}p$$

Set arbitrary normalization: $1^T p = n$ so that

$$p = \left( (1 - \lambda)11^T/n + \lambda LD_c^{-1} \right) p$$

$$=: A p$$

$p$ is an eigenvector of $A$ with eigenvalue 1! And by the properties stochastic matrices, it corresponds to the largest eigenvalue
Google PageRank of pages given by:

\[ p = (1 - \lambda)1 + \lambda LD^{-1}c p \]

Set arbitrary normalization: \( 1^T p = n \) so that

\[ p = ((1 - \lambda)11^T/n + \lambda LD^{-1}c) p \]

\[ =: A p \]

\( p \) is an eigenvector of \( A \) with eigenvalue 1! And by the properties stochastic matrices, it corresponds to the largest eigenvalue

Solve using power method:

\[ p_{k+1} = \frac{A p_k}{1^T A p_k/n} \]

\( p_0 \sim \text{uniform}([0, 1]^n) \)
Matrix completion

Given historical data on how users rated movies in past:

17,700 movies, 480,189 users, 99,072,112 ratings

(Sparsity: 1.2%)

Predict how the same users will rate movies in the future (for $1 million prize)

\[ X = UV^T \]

\[ \mathbf{U} \in \mathbb{R}^{n \times d}, \quad \mathbf{V} \in \mathbb{R}^{m \times d} \]

User \( i \) is assigned a vector for \( \mathbf{u}_i \in \mathbb{R}^d \)

Movie \( j \) is assigned \( \mathbf{v}_j \in \mathbb{R}^d \)

User \( i \) rates movie \( j \) as \( x_{ij} = \mathbf{u}_i^T \mathbf{v}_j \)
Matrix completion

n movies, m users, |S| ratings

$$\arg \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}} \sum_{(i,j,s) \in S} \| (UV^T)_{i,j} - s_{i,j} \|^2_2$$

How do we solve it? With full information?
Matrix completion

n movies, m users, |S| ratings

$$\arg \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}} \sum_{(i,j,s) \in S} \| (UV^T)_{i,j} - s_{i,j} \|^2_2$$
Random projections

PCA finds a low-dimensional representation that reduces population variance

\[
\min_{V_q} \sum_{i=1}^{N} \| (x_i - \bar{x}) - V_q V_q^T (x_i - \bar{x}) \|^2.
\]

\(V_q V_q^T\) is a projection matrix that minimizes error in basis of size \(q\)

\(V_q\) are the first \(q\) eigenvectors of \(\Sigma\)

\[
\Sigma := \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T
\]

But what if I care about the reconstruction of the individual points?

\[
\min_{W_q} \max_{i=1,...,n} \| (x_i - \bar{x}) - W_q W_q^T (x_i - \bar{x}) \|^2
\]
Random projections

$$\min_{W_q} \max_{i=1, \ldots, n} \| (x_i - \bar{x}) - W_q W_q^T (x_i - \bar{x}) \|^2$$

Johnson-Lindenstrauss (1983)

**Theorem 1.1.** (Johnson-Lindenstrauss) Let $\epsilon \in (0, 1/2)$. Let $Q \subset \mathbb{R}^d$ be a set of $n$ points and $k = \frac{20 \log n}{\epsilon^2}$. There exists a Lipschitz mapping $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that for all $u, v \in Q$:

$$(1 - \epsilon)\|u - v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2$$

(independent of $d$)
Random projections

\[
\min_{W_q} \max_{i=1,\ldots,n} \| (x_i - \bar{x}) - W_q W_q^T (x_i - \bar{x}) \|_2
\]

Johnson-Lindenstrauss (1983)

**Theorem 1.1.** (Johnson-Lindenstrauss) Let \( \varepsilon \in (0, 1/2) \). Let \( Q \subset \mathbb{R}^d \) be a set of \( n \) points and \( k = \frac{20 \log n}{\varepsilon^2} \). There exists a Lipschitz mapping \( f : \mathbb{R}^d \to \mathbb{R}^k \) such that for all \( u, v \in Q \):

\[
(1 - \varepsilon) \| u - v \|^2 \leq \| f(u) - f(v) \|^2 \leq (1 + \varepsilon) \| u - v \|^2
\]

**Theorem 1.2.** (Norm preservation) Let \( x \in \mathbb{R}^d \). Assume that the entries in \( A \subset \mathbb{R}^{k \times d} \) are sampled independently from \( N(0, 1) \). Then,

\[
\Pr((1 - \varepsilon) \| x \|^2 \leq \| \frac{1}{\sqrt{k}} Ax \|^2 \leq (1 + \varepsilon) \| x \|^2) \geq 1 - 2e^{-(\varepsilon^2 - \varepsilon^3)k/4}
\]
Other matrix factorizations

\[
X = UV^T
\]

Singular value decomposition

Elements of \(U, S, V\) in \(\mathbb{R}\)

Nonnegative matrix factorization (NMF)

Elements of \(U, S, V\) in \(\mathbb{R}_+\)