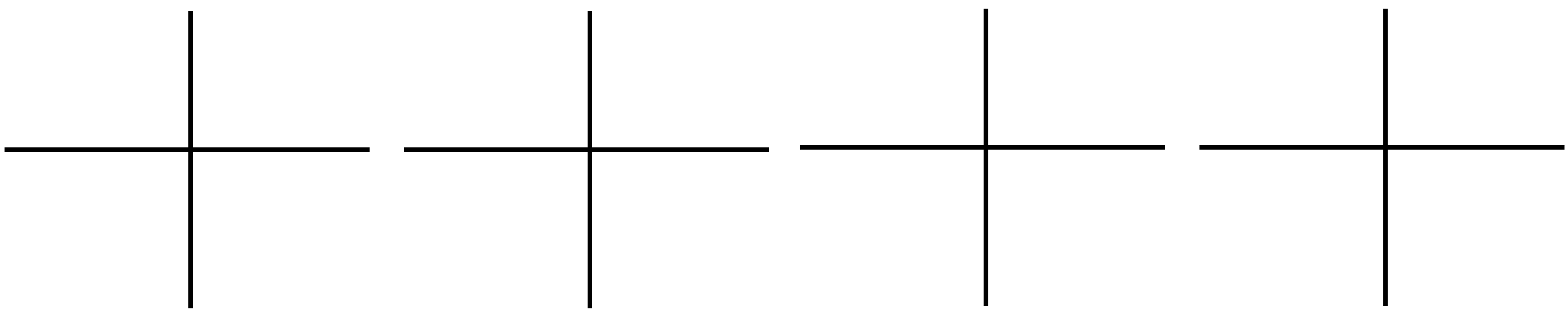
Announcements

- Milestone due tonight
- Fill in the missing plots: $\mathbf{U}, \mathbf{S}, \mathbf{V} = \mathrm{svd}(\mathbf{X})$

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \qquad \mathbf{J} = I - \mathbf{1}\mathbf{1}^T/n$$

$\mathbf{X}$ $\quad$ $\mathbf{J}\mathbf{X}$ $\quad$ $\mathbf{J}\mathbf{X}\mathbf{V}\mathbf{S}^{-1}$ $\quad$ $\mathbf{J}\mathbf{X}\mathbf{V}\mathbf{S}^{-1}\mathbf{V}^T$



Principal Component Analysis (continued)

Machine Learning – CSE546

Kevin Jamieson
University of Washington

November 13, 2017

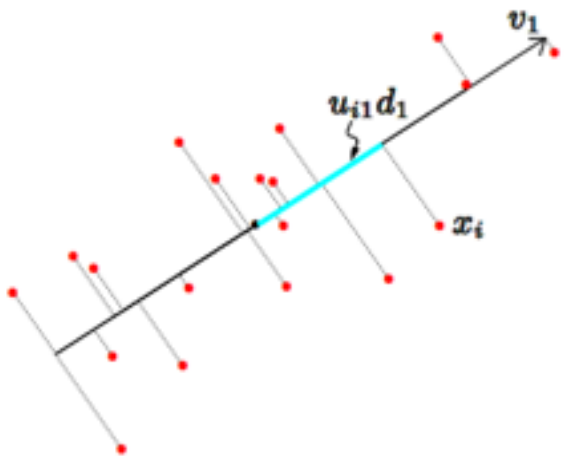
Linear projections

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

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

where $\mathbf{V}_q = [v_1, v_2, \ldots, v_q]$ is orthonormal:

$$\mathbf{V}_q^T \mathbf{V}_q = I_q$$



$\mathbf{V}_q$ are the first q eigenvectors of Σ

$\mathbf{V}_q$ are the first q *principal components*

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

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

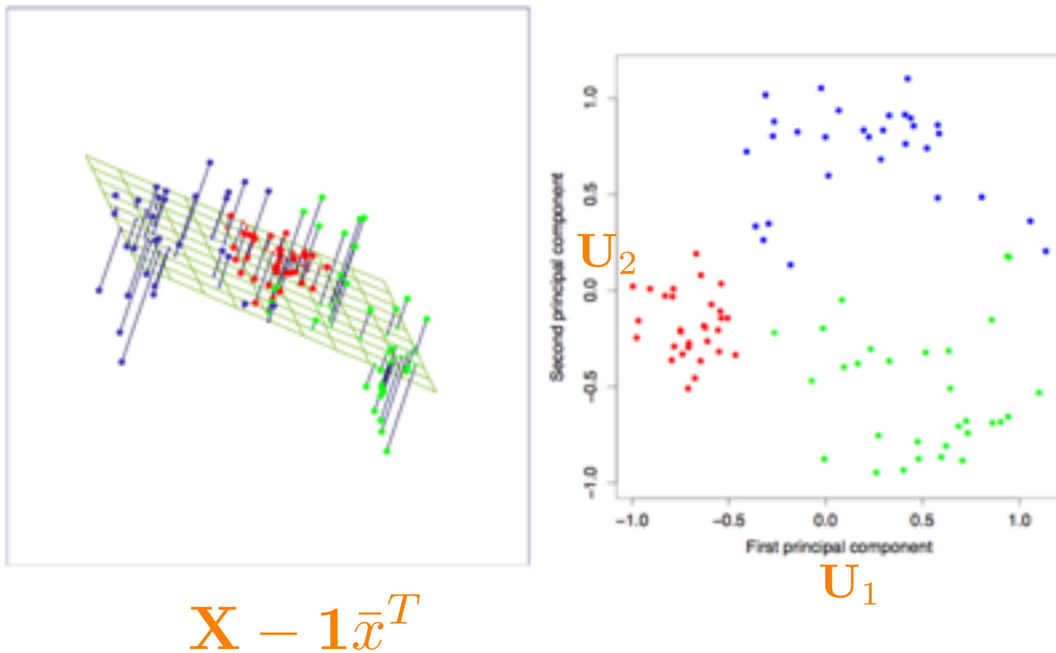
$$(\mathbf{X} - \mathbf{1}\bar{x}^T)\mathbf{V}_q = \mathbf{U}_q \mathrm{diag}(d_1, \ldots, d_q) \qquad \mathbf{U}_q^T \mathbf{U}_q = I_q$$

Singular Value Decomposition defined as

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

Dimensionality reduction

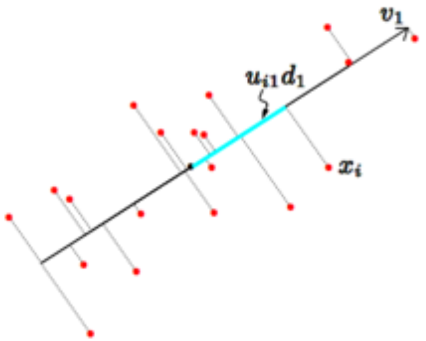
$\mathbf{V}_q$ are the first q eigenvectors of Σ and *SVD* $\mathbf{X} - \mathbf{1}\bar{x}^T = \mathbf{U}\mathbf{S}\mathbf{V}^T$



©Kevin Jamieson 2017

Power method - one at a time

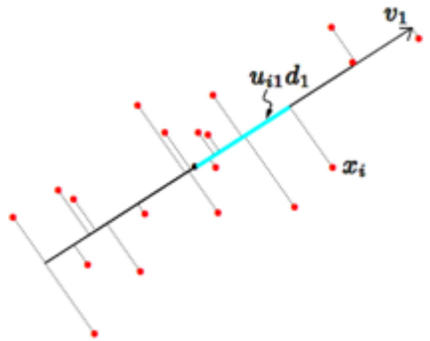
$$\Sigma := \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \qquad v_* = \arg\max_v \, v^T \Sigma v$$



Power method - one at a time

$$\Sigma := \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \qquad v_* = \arg\max_v \; v^T \Sigma v$$

$$v_{k+1} = \frac{\Sigma v_k}{||\Sigma v_k||} \qquad v_0 \sim \mathcal{N}(0, I)$$



Markov chains - PageRank

Markov chains - PageRank

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

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Google PageRank of page i:

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

Markov chains - PageRank

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

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

Google PageRank of pages given by:

$$\mathbf{p} = (1-\lambda)\mathbf{1} + \lambda \mathbf{L}\mathbf{D}_c^{-1}\mathbf{p}$$

Markov chains - PageRank

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

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Google PageRank of pages given by:

$$\mathbf{p} = (1-\lambda)\mathbf{1} + \lambda \mathbf{L}\mathbf{D}_c^{-1}\mathbf{p}$$

Set arbitrary normalization: $\mathbf{1}^T\mathbf{p} = n$ so that

$$\mathbf{p} = \left((1-\lambda)\mathbf{1}\mathbf{1}^T/n + \lambda \mathbf{L}\mathbf{D}_c^{-1}\right)\mathbf{p}$$

$$=: \mathbf{A}\mathbf{p}$$

Markov chains - PageRank

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

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Google PageRank of pages given by:

$$\mathbf{p} = (1-\lambda)\mathbf{1} + \lambda \mathbf{L}\mathbf{D}_c^{-1}\mathbf{p}$$

Set arbitrary normalization: $\mathbf{1}^T\mathbf{p} = n$ so that

$$\mathbf{p} = \left((1-\lambda)\mathbf{1}\mathbf{1}^T/n + \lambda \mathbf{L}\mathbf{D}_c^{-1}\right)\mathbf{p}$$

$$=: \mathbf{A}\mathbf{p}$$

$\mathbf{p}$ is an eigenvector of $\mathbf{A}$ with eigenvalue 1! And by the properties stochastic matrices, it corresponds to the *largest* eigenvalue

Markov chains - PageRank

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

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Google PageRank of pages given by:

$$\mathbf{p} = (1-\lambda)\mathbf{1} + \lambda \mathbf{L}\mathbf{D}_c^{-1}\mathbf{p}$$

Set arbitrary normalization: $\mathbf{1}^T\mathbf{p} = n$ so that

$$\mathbf{p} = \left((1-\lambda)\mathbf{1}\mathbf{1}^T/n + \lambda \mathbf{L}\mathbf{D}_c^{-1}\right)\mathbf{p}$$

$$=: \mathbf{A}\mathbf{p}$$

$\mathbf{p}$ is an eigenvector of $\mathbf{A}$ with eigenvalue 1! And by the properties stochastic matrices, it corresponds to the *largest* eigenvalue

Solve using power method: $\mathbf{p}_{k+1} = \frac{\mathbf{A}\mathbf{p}_k}{\mathbf{1}^T\mathbf{A}\mathbf{p}_k/n}$ $\quad \mathbf{p}_0 \sim \text{uniform}([0,1]^n)$

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)

Alice	1	?	?	4	?	
Bob	?	2	5	?	?	
Carol	?	?	4	5	?	
Dave	5	?	?	?	4	
⋮						

Matrix completion

n movies, m users, |*S*| ratings

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

How do we solve it? With full information?

Matrix completion

n movies, m users, $|S|$ ratings

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

Random projections

PCA finds a low-dimensional representation that reduces population variance

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

$\mathbf{V}_q \mathbf{V}_q^T$ is a *projection matrix* that minimizes error in basis of size q

$\mathbf{V}_q$ are the first q eigenvectors of Σ

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

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

$$\min_{\mathbf{W}_q} \max_{i=1,\dots,n} ||(x_i - \bar{x}) - \mathbf{W}_q \mathbf{W}_q^T (x_i - \bar{x})||^2$$

Random projections

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

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 Lipshcitz mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in Q$:*

(independent of d)

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

Random projections

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

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 Lipshcitz mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in Q$:*

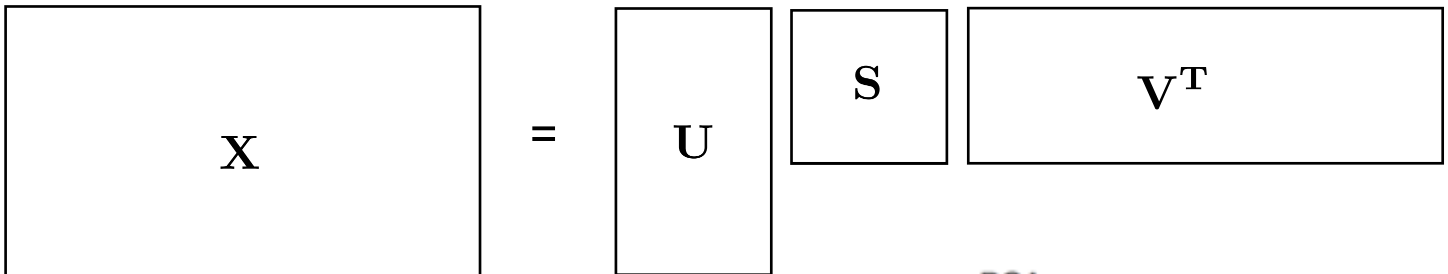
(independent of d)

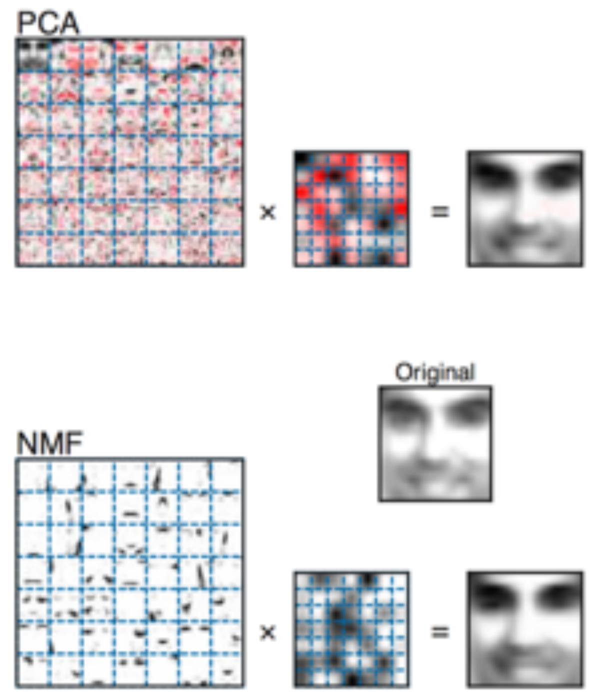
$$(1-\epsilon)\|u-v\|^2 \le \|f(u) - f(v)\|^2 \le (1+\epsilon)\|u-v\|^2$$

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,*

$$\Pr((1-\epsilon)\|x\|^2 \le \|\frac{1}{\sqrt{k}} Ax\|^2 \le (1+\epsilon)\|x\|^2) \ge 1 - 2e^{-(\epsilon^2 - \epsilon^3)k/4}$$

Other matrix factorizations





Singular value decomposition

Elements of $\mathbf{U}, \mathbf{S}, \mathbf{V}$ in $\mathbb{R}$

Nonnegative matrix factorization (NMF)

Elements of $\mathbf{U}, \mathbf{S}, \mathbf{V}$ in $\mathbb{R}_+$