Midterm feedback

- Welcome back!
- OSC: Don’t copy HW!!!
- Midterm: returned today

Survey results

- Consistent notation; correct quiz answers
- More depth (math and application); more intuition & derivation
  - Recitation as review; no new material
  - Too much math in HW; too little in lecture
- Slides and lecture notes more complete
- Too fast
- HW: too much; not clear enough; too many errors
  - Autograder; Output shape for programming problems
  - Ask for more explanation
- Faster piazza response time

Unsupervised Learning

- **Spectral methods**
  - Eigenvector/singular vector decomposition (SVD)
  - PCA, CCA
- **Reconstruction methods**
  - PCA, ICA, auto-encoders
- **Clustering and Probabilistic methods**
  - K-means
  - Gaussian mixtures
  - Latent Dirichlet Allocation (LDA)

SVD

Lyle Ungar

Learning objectives
SVD and “thin SVD”
Matrix norms
Generalized inverses

Eigenvectors (review)

- **$A v_i = \lambda_i v_i$**
- **Eigen-decomposition of a symmetric matrix A** (n x n)
  - **A** = **VDV^T**
- **V: orthogonal, $V^TV=I$** (n x n)
  - Columns of V are the *eigenvectors of A*
- **D: diagonal** (n x n)
  - Diagonal elements of D are the *eigenvalues of A*
  - All non-negative if $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X}$
  - Reported in *decreasing* order of magnitude down the diagonal

We don't compute eigenvectors

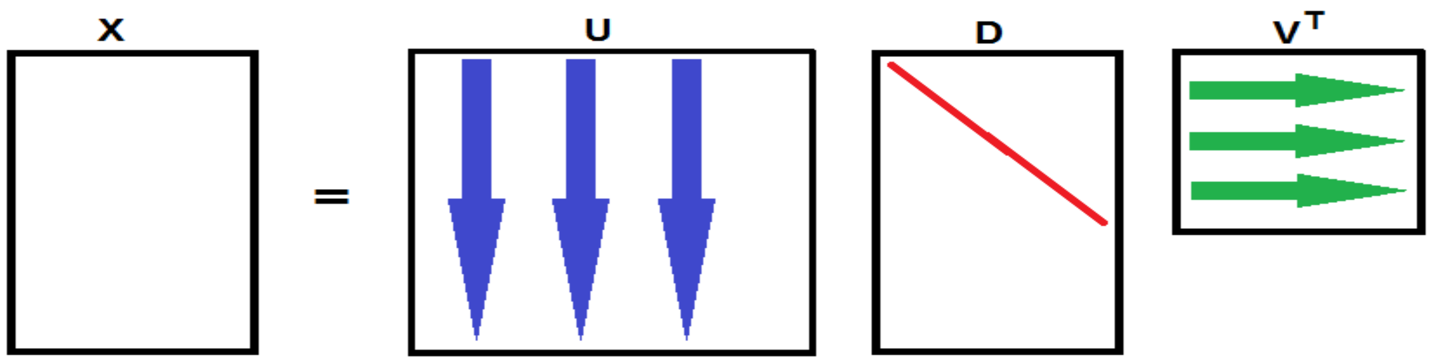
- **What symmetric matrix have we seen?**
- **In practice we rarely compute eigenvectors**
  - Why not?

Singular Value Decomposition

- **Singular value decomposition of matrix X** (n x p)
  - **X** = **UDV**^{T}
- **U: orthogonal, U^{T}U=I** (n x n)
  - Columns of **U** are the *left singular vectors of **X***
- **D: diagonal** (n x p)
  - Diagonal elements of **D** are the *singular values of **X***
- **V: orthogonal, V^{T}V=I** (p x p)
  - Columns of **V** are the *right singular vectors of **X****

SVD

Singular value decomposition of X: **X = UDV^T**



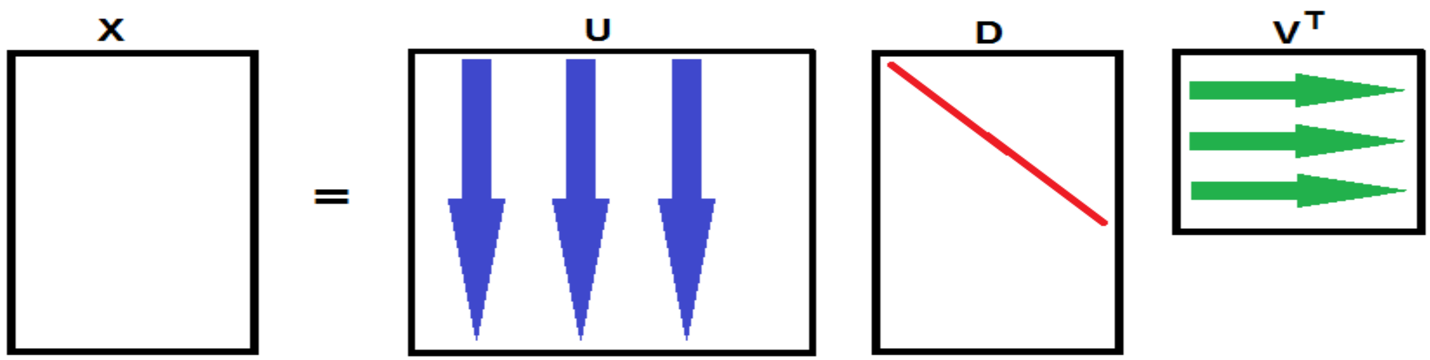
Let k = min(n,p). Then: $\mathbf{X} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{k}} D_{ii}\boldsymbol{u}_i\boldsymbol{v}_i^T$

Since all $\boldsymbol{u}_i, \boldsymbol{v}_i$ are unit vectors, the importance of the i'th term in the sum is determined by the size of D_{ii}.

- $X_{n*p} = U\ D\ V^T$
- What are the dimensions of U D and V?
- What are the eigenvectors of X^TX?
- What are the eigenvalues of X^TX?

Thin SVD – pick a smaller k

Singular value decomposition of X: **X = UDV^T**



Let k = min(n,p). Then: $\mathbf{X} = \sum_{i=1}^{k} D_{ii}\boldsymbol{u}_i\boldsymbol{v}_i^T$

Since all $\boldsymbol{u}_i, \boldsymbol{v}_i$ are unit vectors, the importance of the i'th term in the sum is determined by the size of D_{ii}.

SVD and eigenvalues/eigenvectors

$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\boldsymbol{T}},$ $\quad\quad$ $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} = \boldsymbol{V}\left(\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D}\right)\boldsymbol{V}^{\boldsymbol{T}}$

The columns $\mathbf{v_1}, \ldots \mathbf{v_p}$ of **V** are the *eigenvectors* of the covariance matrix $\mathbf{X^T X}$**.** Hence we can write

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{p}} (D_{ii})^2\, \boldsymbol{v}_i \boldsymbol{v}_i^T$$

From before:

$$\boldsymbol{X} = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{k}} D_{ii} \boldsymbol{u}_i \boldsymbol{v}_i^T$$

k = min(n,p).

D_{ii} are singular values of X, $(D_{ii})^2$ are eigenvalues of $\mathbf{X^T X}$

Frobenius norm

- How to measure the size of a matrix?

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\operatorname{trace}(A^{\dagger}A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)}$$

- Where σ_i are the singular values.
- One can also use an L_1 norm $||A||_1 = ||\sigma||_1$

Generalized Inverses

- Linear regression estimates w in $y = Xw$
- This uses a pseudo-inverse ("Moore-Penrose inverse") X^+ of X, so
  - $w = X^+ y$
- Thus far, we have done this by
  - $X^+ = (X^T X)^{-1} X^T$

Generalized Inverses

- We can also compute inverses using SVD
- The idea:

$$X^{+} = (U\Lambda^{-1}V^{T})^{T} = V\Lambda^{-1}U^{T}$$

- You can't take the inverse of a rectangular matrix, but we can approximate it using the thin SVD

$$X^{+} = V_k\Lambda_k^{-1}U_k^{T}$$

Pseudo-inverse of $X = U D V^T$

- **What are the dimensions of $X^+ = V D^{-1} U^T$**
- **What is $X X_k^+$**
  - $X X^+ = U D V^T V D^{-1} U^T$

Power Method

- **Power method for a square matrix A**
  - Write any $\mathbf{x} = \Sigma_i z_i \mathbf{v}_i$ where $z_i = \mathbf{v}_i^T\mathbf{x}$
  - Then $Ax = A \Sigma_i z_i \mathbf{v}_i = \Sigma_i z_i A \mathbf{v}_i = \Sigma_i z_i \lambda_i \mathbf{v}_i$
  - So $AAAAx = A^4x = = \Sigma_i z_i \lambda_i^4 \mathbf{v}_i$
- **Find the largest eigenvalue/eigenvector**
  - Project it off from x and repeat
    - $\mathbf{x} := \mathbf{x} - (\mathbf{v}_1^T\mathbf{x})\, \mathbf{x}$

Fast 'Randomized' SVD

- **Generalizes the power method**
- **Input:**
  - matrix **A** of size n × p,
  - the desired hidden state dimension k,
  - the number of "extra" singular vectors, l
- **Simultaneously find all the largest singular values/vectors by alternately left and right multiplying by A**

Randomized SVD

1. Generate a $(k+l) \times n$ random matrix Ω

2. Find the SVD $U_1 D_1 V_1^T$ of ΩA, and keep the $k+l$ components of V_1 with the largest singular values

3. Find the SVD $U_2 D_2 V_2^T$ of AV_1, and keep the 'largest' $k+l$ components of U_2

4. Find the SVD $U_3 D_3 V_{final}^T$ of $U_2^T A$, and keep the 'largest' k components of V_{final}

5. Find the SVD $U_{final} D_4 V_4^T$ of AV_{final} and keep the 'largest' k components of U_{final}

Output: The left and right singular vectors U_{final}, V_{final}^T

You are not required to know this

What you should know

- Eigenvalues/vectors & singular values/vectors
- Eigenvectors as a basis
- Thin SVD
- Frobenius norm
- Pseudo (“Moore-Penrose”) inverse
- Power method

- **What is an efficient way to do linear regression?**
  - $\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T\mathbf{y}$
  - How does it scale with n and p?