Feedback

- THE COURSE MOVES TOO FAST!!
  - Too much material
  - Profs talk too fast
- Good breadth
- More math rigor
- Less math – stick to intuition
- More coding
  - And fewer autograder problems
  - Change to python
Unsupervised learning

- **Spectral methods**
  - Eigenvector/singular vector decomposition
  - PCA/CCA

- **Reconstruction methods**
  - PCA/ICA/autoencoders

- **Clustering and Probabilistic methods**
  - K-means
  - Gaussian mixtures
  - Latent Dirichlet Allocation (LDA)
SVD, PCA and PCR

Lyle Ungar

Based in part on slides by Jia Li (PSU) and Barry Slaff (Upenn)
Eigenvectors (review)

- $A \mathbf{v}_i = \lambda_i \mathbf{v}_i$
- Eigen-decomposition of a symmetric matrix $A$ $(n \times n)$
  - $A = \mathbf{VDV}^T$
- $\mathbf{V}$: orthogonal, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ $(n \times n)$
  - Columns of $\mathbf{V}$ are the eigenvectors of $A$.
- $\mathbf{D}$: diagonal $(n \times n)$
  - Diagonal elements of $\mathbf{D}$ are the eigenvalues of $A$.
  - All non-negative
  - In decreasing order of magnitude down the diagonal.
(no) 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, $U^TU=I$ (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, $V^TV=I$ (p x p)
  - Columns of $V$ are the *right singular vectors of $X$.*
SVD

Singular value decomposition of \( X \): \( X = U D V^T \)

Let \( k = \min(n,p) \). Then:
\[
X = \sum_{i=1}^{k} D_{ii} u_i v_i^T
\]

Since all \( u_i, v_i \) are unit vectors, the importance of the \( i \)'th term in the sum is determined by the size of \( D_{ii} \).
Thin SVD – pick a smaller $k$

Singular value decomposition of $X$: $X = U D V^T$

Let $k = \min(n,p)$. Then: $X = \sum_{i=1}^{k} D_{ii} u_i v_i^T$

Since all $u_i, v_i$ 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

\[ X = UDV^T, \quad X^TX = V(D^TD)V^T \]

The columns \( v_1, \ldots, v_p \) of \( V \) are the eigen vectors of the covariance matrix \( X^TX \). Hence we can write

\[ X^TX = \sum_{i=1}^{p} (D_{ii})^2 \, v_i v_i^T \]

From before:

\[ X = \sum_{i=1}^{k} D_{ii} u_i v_i^T \]

\( k = \min(n,p) \).

\( D_{ii} \) are singular values of \( X \), \( (D_{ii})^2 \) are eigenvalues of \( X^TX \).
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
  - \( X^+ = (X^TX)^{-1} X^T \)
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.
  \[
  X^+ = V_k\Lambda_k^{-1}U_k^T
  \]
Power Method

Power method for a square matrix $A$

- Write any $x = \Sigma_i z_i \cdot v_i$ where $z_i = v_i \cdot x$
- Then $Ax = A \Sigma_i z_i \cdot v_i = \Sigma_i z_i \cdot A \cdot v_i = \Sigma_i z_i \cdot \lambda_i \cdot v_i$
- So $AAAAx = A^4x = \Sigma_i z_i \cdot \lambda_i^4 \cdot v_i$

Find the largest eigenvalue/eigenvector

- Project it off from $x$ and repeat

- $x := x - v_1 \cdot x$
Fast SVD

- Generalizes the power method
- Input:
  - matrix $A$ of size $n \times p$,
  - 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 $A$
Fast SVD

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

2. Find the SVD \(U_1D_1V_1^T\) of \(\Omega A\), and keep the \(k + l\) components of \(V_1\) with the largest singular values.

3. Find the SVD \(U_2D_2V_2^T\) of \(AV_1\), and keep the ‘largest’ \(k + l\) components of \(U_2\).

4. Find the SVD \(U_3D_3V_{final}^T\) of \(U_2^TA\), and keep the ‘largest’ \(k\) components of \(V_{final}\).

5. Find the SVD \(U_{final}D_4V_4^T\) of \(AV_{final}\) and keep the ‘largest’ \(k\) components of \(U_{final}\).

Output: The left and right singular vectors \(U_{final}, V_{final}^T\).
What you should know

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