Lecture 8

Announcements

- HW3 + Labs updated

Readings

Strang I.9, I.10, I.11

Outline

- Proof of Eckart-Young in Frobenius
- Rayleigh quotients
- Fisher's Linear Discriminant Analysis.

Proof of Eckart-Young in Frobenius

Review: Frobenius

① $\|A\|_F^2 = |a_{11}|^2 + \dots + |a_{m1}|^2 + |a_{12}|^2 + \dots + |a_{mn}|^2$ (treat A as an mn by 1 vector)

② $\|A\|_F^2 = \text{trace of } A^TA$ (diagonal entries of A^TA contain ℓ^2 norms of columns of A)

③ $\|A\|_F^2 = \sigma_1^2 + \dots + \sigma_r^2$ $\left(\text{sum of eigenvalues of } A^TA = \text{trace}(A^TA)\right)$

So, if $A = U\Sigma V^T$, $\|A\|_F^2 = \|\Sigma\|_F^2$

Eckart-Young

If $A = U\Sigma V^T$, then $A_k = \sigma_1 \vec{u}_1 \vec{v}_1^T + \dots + \sigma_k \vec{u}_k \vec{v}_k^T$ is the best rank-k approximation to A. That is, if B has rank k, then

$$\|A - A_k\|_F \leq \|A - B\|_F$$

Proof Trick: Take SVD of B, not A!

Suppose B is the closest rank-k matrix to A

Take SVD of B

$$B = U \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} V^T$$

where D is k by k diagonal matrix

$$A = U \underbrace{\begin{bmatrix} L+E+R & F \\ G & H \end{bmatrix}}_{U^TAV} V^T$$

where L is strictly lower triangular
R is strictly upper triangular
E is diagonal

Consider a rank $\leq k$ matrix C:

$$C = U \begin{bmatrix} L+D+R & F \\ 0 & 0 \end{bmatrix} V^T$$

$$\|A-B\|_F^2 = \|L+E+R-D\|_F^2 + \|F\|_F^2 + \|G\|_F^2 + \|H\|_F^2$$

$$= \|L\|_F^2 + \|E-D\|_F^2 + \|R\|_F^2 + \|F\|_F^2 + \|G\|_F^2 + \|H\|_F^2$$

$$\|A-C\|_F^2 = \|E-D\|_F^2 + \|G\|_F^2 + \|H\|_F^2$$

$$\|A-B\|_F^2 = \|A-C\|_F^2 + \|L\|_F^2 + \|R\|_F^2 + \|F\|_F^2$$

Set $L, R, F = 0$

Analogously, we can show $G = 0$

$$A = U\begin{bmatrix} E & 0 \\ 0 & H \end{bmatrix} V^T \qquad B = U\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} V^T$$

Since B is closest to A, $D = E$

Notes:

- Singular values of H are the $r-k$ smallest singular values of A.
- The error $\|A-B\|_F = \|H\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}$

Rayleigh Quotients

Another way to understand SVD:

if we maximize $\frac{\|A\vec{x}\|}{\|\vec{x}\|}$, the maximum is σ_1 at $\vec{x} = \vec{v}_1$

How do we derive this?

maximize $\frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} = \frac{\vec{x}^T A^T A \vec{x}}{\vec{x}^T\vec{x}} = \boxed{\frac{\vec{x}^T S \vec{x}}{\vec{x}^T\vec{x}}}$ (Rayleigh quotient)

$$\frac{\partial}{\partial x_i}(\vec{x}^T\vec{x}) = \frac{\partial}{\partial x_i}\left(x_1^2 + \dots + x_n^2\right) = 2x_i$$

$$\frac{\partial}{\partial x_i}(\vec{x}^T S\vec{x}) = \frac{\partial}{\partial x_i}\left(\sum_i\sum_j S_{ij}\, x_i x_j\right) = 2\sum_j S_{ij} x_j = 2(S\vec{x})_i$$

apply quotient rule

$$2(\vec{x}^T\vec{x})(S\vec{x})_i - 2(\vec{x}^T S\vec{x})\, x_i = 0$$

This is satisfied when $(Sx)_i = \lambda x_i$ for all i

so when $S\vec{x} = \lambda\vec{x}$

The maximum is at $\vec{v}_1$ (first right singular vector of A)

<u>def</u> The Rayleigh quotient for a symmetric matrix S is a function

$$R : \mathbb{R}^n - \{\vec{0}\} \longrightarrow \mathbb{R}$$

$$R(\vec{x}) = \frac{\vec{x}^T S\vec{x}}{\vec{x}^T\vec{x}}$$

<u>note</u> $R(k\vec{x}) = R(\vec{x})$ (scaling invariant)

So we can focus on the unit sphere $S_n = \{\vec{x}\in\mathbb{R}^n : \|\vec{x}\| = 1\}$

The Rayleigh quotient is essentially a quadratic form over the unit sphere:

<u>ex</u> $S = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$ $\quad R(\vec{x}) = \dfrac{x_1^2 + 2x_2^2 + 6x_1x_2}{x_1^2 + x_2^2}$ $\quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

<u>note</u> The max value of R is the largest eigenvalue of S.

The min value is the smallest eigenvalue.

Generalized eigenvalues + eigenvectors

Let M be a symmetric matrix.

<u>def</u> Generalized Rayleigh quotient $R(\vec{x}) = \dfrac{\vec{x}^T S \vec{x}}{\vec{x}^T M \vec{x}}$

when we work with generalized eigenvalues + eigenvectors

instead of solutions of $S\vec{x} = \lambda \vec{x}$

we find solutions of $S\vec{x} = \lambda M \vec{x}$

If M is positive definite, then $\max R(\vec{x})$ is equal to the largest eigenvalue of $M^{-1}S$.

$M^{-1}S$ may not be symmetric, but $M^{-1/2} S M^{-1/2}$ is symmetric

Importantly: $M^{-1}S$ and $M^{-1/2} S M^{-1/2}$ have the same eigenvalues.

So: we want to find the eigenvalues of $H = M^{-1/2} S M^{-1/2}$

Solving $S\vec{x} = \lambda M \vec{x}$ is equivalent to maximizing $\dfrac{\vec{y}^T H \vec{y}}{\vec{y}^T \vec{y}} = \dfrac{\vec{x}^T S \vec{x}}{\vec{x}^T M \vec{x}}$

where $\vec{x} = M^{-1/2} \vec{y}$

$\longrightarrow$ so apply typical strategies to find λ

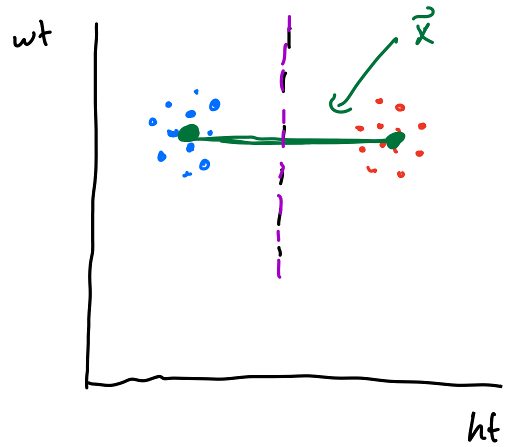
Fisher's linear discriminant analysis

data

$$X_1 = \begin{array}{c} \begin{array}{cc} ht & wt \end{array} \\ \begin{bmatrix} & \\ & \end{bmatrix} \\ n_1 \times 2 \end{array} \qquad X_2 = \begin{array}{c} \begin{array}{cc} ht & wt \end{array} \\ \begin{bmatrix} & \\ & \end{bmatrix} \\ n_2 \times 2 \end{array}$$

Σ_1 Σ_2

sample covariance



Idea: identify line/plane/hyperplane that best separates two groups.

$m_1 = \begin{bmatrix} \leftarrow \text{mean ht for pop 1} \\ \leftarrow \text{mean wt for pop 1} \end{bmatrix}$ $\quad m_2 = \begin{bmatrix} \\ \end{bmatrix}$

$2 \times 1 \qquad\qquad 2 \times 1$

Fisher's LDA maximizes the separation ratio:

$$R(\vec{x}) = \frac{(\vec{x}^T \vec{m}_1 - \vec{x}^T \vec{m}_2)^2}{\vec{x}^T \Sigma_1 \vec{x} + \vec{x}^T \Sigma_2 \vec{x}} = \frac{(\vec{x}^T(\vec{m}_1 - \vec{m}_2))^2}{\vec{x}^T(\Sigma_1 + \Sigma_2)\vec{x}} = \frac{\vec{x}^T S \vec{x}}{\vec{x}^T M \vec{x}}$$

$S = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T$

$M = (\Sigma_1 + \Sigma_2)$

The maximizing vector will be $\boxed{\vec{v} = M^{-1}(\vec{m}_1 - \vec{m}_2)}$