Lecture 6 Matrix Perturbations

Monday, September 12, 2016 2:02 PM

- Whitening and Canonical Correlation Analysis.

  - Now consider covariance matrix
  $$M = \sum_{i=1}^{m} x_i x_i^T$$

  - Often want to perform linear transformation
  $$z_i = W^T x_i \quad \text{such that}$$
  $$M_z = \sum_{i=1}^{m} z_i z_i^T = I$$
  - Motivation: result invariant to linear transformation.

  - How to find W?

  SVD of M: $M = UDU^T$
  $$W = UD^{-\frac{1}{2}}$$
  $$M_z = \sum_{i=1}^{m} (W^T x_i)(W^T x_i)^T = W^T M W$$
  $$= D^{-\frac{1}{2}} U^T (U^T D U) U D^{-\frac{1}{2}}$$
  $$= I$$
  Whitening matrix is not unique (RW also works when $RR^T = I$)

  - Canonical Correlation analysis
    Given data (x_1, y_1) (x_2, y_2) $\cdots$ (x_m, y_m)
    want to find direction (u, v) such that
    (x, y) are <u>most</u> correlated in direction (u, v).

  - More precisely: find u, v such that
  $$\mathbb{E}[\langle u, x\rangle^2] = \mathbb{E}[\langle v, y\rangle^2] = 1$$
  $$\mathbb{E}[\langle u, x\rangle\langle v, y\rangle] \text{ maximized.}$$
  - Solve using SVD
    Let $M_x = \mathbb{E}[xx^T]$, $M_y = \mathbb{E}[yy^T]$
    $$M_{xy} = \mathbb{E}[xy^T]$$
    Let W_x, W_y be whitening matrix of M_x, M_y.
  Claim: If $\|\bar{u}\| = 1$, then $u = W_x \bar{u}$ satisfies

$$u^T M_x u = 1.$$

Proof: just by $W_x^T M_x W_x = I$.

Therefore, we are trying to find $\|\bar{u}\| = \|\bar{v}\| = 1$ s.t.

$$\mathbb{E}[\langle W_x \bar{u}, x\rangle \langle W_y \bar{v}, y\rangle] \text{ is maximized}$$

but $\mathbb{E}[\langle W_x \bar{u}, x\rangle \langle W_y \bar{v}, y\rangle]$

$$= \bar{u}^T \mathbb{E}[W_x^T x y^T W_y] \bar{v}$$

$$= \bar{u}\ W_x^T M_{xy} W_y \bar{v}$$

so we can simply do SVD of $W_x^T M_{xy} W_y$

Matrix Perturbations

- Typical Scenario

  "true matrix" A (often low rank)

  noise N

  how does singular value/vectors change from A to $A+N$?

- main intuition: singular vectors are stable if

  $\|N\| \ll$ spectral gap.

- Weyl's Theorem: $\forall i$

$$\sigma_i(A) - \|N\| \leq \sigma_i(A+N) \leq \sigma_i(A) + \|N\|$$

  To get intuition, think of A as diagonal

  (can also do this wlog in proof)

- Wedin's Theorem:

  If $A = [U_1\ U_2] \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix} [V_1\ V_2]^T$

$$A+N = [\hat{U}_1\ \hat{U}_2] \begin{bmatrix} \hat{D}_1 & \\ & \hat{D}_2 \end{bmatrix} [\hat{V}_1\ \hat{V}_2]^T$$

  if $\|N\| < \frac{1}{4} \cdot \min_{i,j} |D_1(i,i) - \hat{D}_2(j,j)|$

if $\|N\| < \frac{1}{4} \cdot \underbrace{\min_{i,j} \left| D_1(i,i) - \hat{D}_2(j,j) \right|}_{\text{gap } \tau}$

then $\sin\Theta(U_1, \hat{U}_1) \le \frac{4\|N\|}{\tau}$.

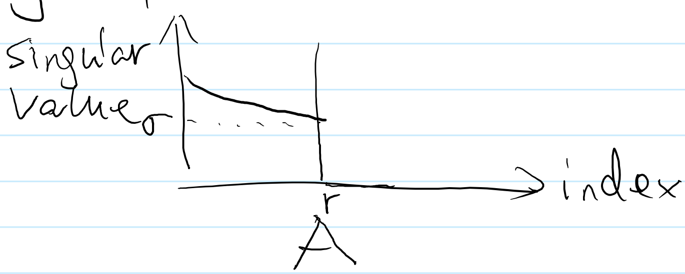
Def: $\Theta(U_1, \hat{U}_1) = \max_{u_1 \in U_1} \min_{\hat{u}_1 \in \hat{U}_1} \theta(u_1, \hat{u}_1)$

"principal angle between two subspace"

$$\sin\Theta(U_1, \hat{U}_1) = \|(I - U_1 U_1^T)\hat{U}_1\| = \|(I - \hat{U}_1 \hat{U}_1^T) U_1\|$$
$$= \|U_1 U_1^T - \hat{U}_1 \hat{U}_1^T\|$$

Example: A is rank r, $\sigma_r(A) = \sigma \gg \|N\|$

Weyl's Theorem



Wedin's Theorem: $\min D_1 = \sigma$ $\max \hat{D}_2 \le \|N\|$

Gap $\tau \ge \sigma - \|N\|$

$$\sin\Theta(U_1, \hat{U}_1) \le \frac{4\|N\|}{\sigma - \|N\|}$$

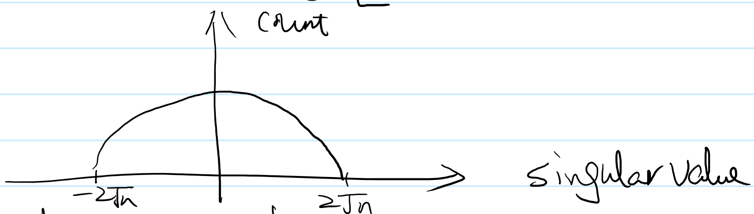
How to prove $\|N\|$ is small?

1. random matrix

for $N \in \mathbb{R}^{n \times n}$, if all entries are iid Gaussian
± 1
...

$\|N\| \le O(\sqrt{n})$ $(2\sqrt{n})$

Wigner's Semicircle Law



histogram of singular values

Limitations: distribution of eigenvalues only known for a few distributions.