Note Title 2024-10-18

Convergence of CG and polynomial approximation:

$$\frac{\|x_n - x_*\|_Q}{\|x_0 - x_*\|_Q} \le \min_{r(t)} \max_{\lambda_1, \dots \lambda_m} |r(\lambda_i)|$$

$r(t)$ varies over polynomials of degree $\le n$ s.t. $r(0)=1$

Proof:

$$\|x_0 - x_*\|_Q^2 = \|-x_*\|_Q^2 = \|x_*\|_Q^2 = x_*^T Q x_*$$

If $Q = UDU^T$ is an eigenvalue decomp., and $x_* = Uc$ are the coordinates of x_* in the eigenvector basis, then

$$x_*^T Q x_* = c^T \cancel{U^T U} D \cancel{U^T U} c = [c_1 \cdots c_m] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$$

$$= \lambda_1 c_1^2 + \lambda_2 c_2^2 + \dots + \lambda_m c_m^2.$$

$\|x_n - x_*\|$ $\qquad x_n = \underset{x \in K_n(Q,v)}{\arg\min} \|x - x_*\|_Q$

$$K_n(Q,v) = \{ p(Q)v : p \text{ polynomial of degree } < n \}$$

$$x_n = \underset{p(t)}{\arg\min} \|p(Q)v - x_*\|_Q = \underset{p(t)}{\arg\min} \|x_* - p(Q)Qx_*\|_Q$$

$$= \underset{r(t)}{\arg\min} \|r(Q)x_*\|_Q$$

where $r(t) = 1 - p(t)t$ varies over all polynomials of degree $\le n$ such that $r(0)=1$

If $Q = UDU^T$, $r(Q) = U \begin{bmatrix} r(\lambda_1) & & & \\ & r(\lambda_2) & & \\ & & \ddots & \\ & & & r(\lambda_m) \end{bmatrix} U^T$

(from earlier results)

$$= \arg\min_{r(t)} \left\| U \begin{bmatrix} r(\lambda_1) & & \\ & \ddots & \\ & & r(\lambda_m) \end{bmatrix} \cancel{U^T U} c \right\|_Q = \arg\min \left\| U \begin{bmatrix} r(\lambda_1) c_1 \\ \vdots \\ r(\lambda_m) c_m \end{bmatrix} \right\|_Q$$

$$= \arg\min_r \sqrt{c_1^2 |r(\lambda_1)|^2 \lambda_1 + c_2^2 |r(\lambda_2)|^2 \lambda_2 + \dots + c_m^2 |r(\lambda_m)|^2 \lambda_m}$$

$$\frac{\|x_n - x_*\|_Q^2}{\|x_0 - x_*\|_Q^2} = \min_{r(t)} \frac{c_1^2 |r(\lambda_1)|^2 \lambda_1 + \dots + c_m^2 |r(\lambda_m)|^2 \lambda_m}{c_1^2 \lambda_1 + \dots + c_m^2 \lambda_m} \leq$$

$$\leq \min_{r(t)} \frac{\max |r(\lambda_i)|^2 \cancel{(c_1^2 \lambda_1 + \dots + c_m^2 \lambda_m)}}{\cancel{c_1^2 \lambda_1 + \dots + c_m^2 \lambda_m}}.$$

Singular value decomposition

$A \in \mathbb{R}^{m \times n}$ $\quad m > n \quad$ $\underset{m\times m}{U}\,\underset{m\times n}{S}\,\underset{n\times n}{V^T}$: $A = USV^T$, with U, V orthogonal,

$$S = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & 0 & \end{bmatrix} \qquad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$$

$$A = u_1 \sigma_1 v_1^T + \dots + u_{\min(m,n)} \sigma_{\min(m,n)} v_{\min(m,n)}^T$$

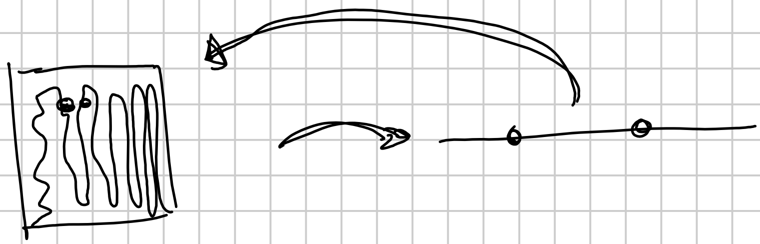
Theorem (Eckhart-Young): the solution of

$$X_k = \arg\min_{\operatorname{rk} X \leq k} \|A - X\|$$

is

$$X_k = u_1 \sigma_1 v_1^T + \dots + u_k \sigma v_k^T$$

$m+n+1$



If we compress one $m \times n$ matrix to rank k,
we save space

$$mn \longrightarrow mk + nk + k$$

	Test A	test B	test C
Student 1	90	95	85
Student 2	32	35	31
Student 3	70	75	71
Student 4	70	31	72

$= A$

If A were rank 1,

$$A = xy^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & \\ \vdots & & \ddots \\ x_4y_1 & \cdots & x_4y_3 \end{bmatrix}$$

result of student i on test j is $x_i y_j$

$$\underbrace{x_i}_{\text{student ability}} \underbrace{y_j}_{\text{test difficulty}}$$

rank 1 $\Leftrightarrow$ perfect structure in your data.

If A has rank 2,

$$A = xy^T + wz^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \begin{bmatrix} y_1 & y_2 + y_3 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}$$

$$A_{ij} = x_i y_j + w_i z_j \qquad x, y, z, w \geq 0$$

Idea:

2 different topics, e.g. programming and mathematics

Each student i gets a score that reflect their abilities in the two subjects (x_i, w_i) and the amount of math difficulty / programming difficulty in each exercise j (y_j, z_j)

---

svd approximation: $A \approx u_1 \sigma_1 v_1^T$ gives the best possible rank-1 approximation of true student scores

Entries of $u_1 \sim$ estimated ranking of students

$v_1 \sim$ est. ranking of problems.

"Best" = minimum $\sum_{i=1}^{m} \sum_{j=1}^{n} (A_{ij} - u_{i1}\sigma_1 v_{j1})^2$

In statistics terms:

model: $\text{Score}_{ij} = x_i y_j + \epsilon_{ij}$

ϵ_{ij} = error $\sim$ Gaussian$(0, \sigma)$

maximum likelihood estimator $\Leftrightarrow$ most likely choice of x, y that could have generated this data.

$$ML \Leftrightarrow \min \sum_{i=1}^{m} \sum_{j=1}^{n} (A_{ij} - x_i y_j)^2 = \min \|A - xy^T\|_F^2$$

$$\begin{array}{cc} & \text{test 1} \;\cdots\; \text{test 3} \\ \begin{array}{c}\text{Student 1}\\ \vdots \\ \text{Student 4}\end{array} & \begin{bmatrix} \quad & \quad & \quad \\ & & \\ & & \end{bmatrix} \end{array}$$

rank 2: $A_{ij} = x_i y_j + w_i z_j$

two different topics, each with ability / difficulty

With an SVD:

$$A_{ij} = U_1 \sigma_1 V_1^T + U_2 \sigma_2 V_2^T + U_3 \sigma_3 V_3^T$$

overall ability

higher rank: further corrections for further topics

Since U_i, V_j have norm 1, the σ_i tell me how large the corrections are.

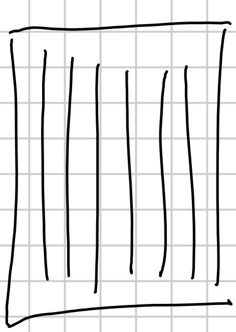
$\sigma_1 \gg \sigma_2$: the rank-2 approximation is only slightly changed w.r.t. the rank 1 approximation

→ the classification with only 1 topic was already pretty good

$\sigma_1 \approx \sigma_2$: classification into 2 topics makes a big change.

Related: $\boxed{\text{latent semantic analysis}}$, principal component analysis (PCA)

PCA: given a data matrix

A = 

x_j = columns of A = observations

$$m = \text{mean} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

$$\hat{x}_j = x_j - m$$

$$\hat{A} = \begin{bmatrix} | & & | \\ \hat{x}_1 & \cdots & \hat{x}_n \\ | & & | \end{bmatrix}.$$

Variances and covariances of each component of the data:

Covariance matrix $C = \frac{1}{n-1}\left(\hat{x}_1 \hat{x}_1^T + \dots + \hat{x}_n \hat{x}_n^T\right)$

principal components = eigenvectors of the covariance matrix

This can also be seen in terms of the SVD:

$$\hat{A} = USV^T$$

$$C = \frac{1}{n-1}\hat{A}\hat{A}^T = \frac{1}{n-1} US\cancel{V^T V}S^T U^T = \frac{1}{n-1} U \begin{bmatrix} \sigma_1^2 & & & & & \\ & \sigma_2^2 & & & & \\ & & \ddots & & & \\ & & & \sigma_n^2 & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix} U^T$$