Board Notes - Reconstruction Error

Monday, April 3, 2017 11:20 AM

$$\| \hat{x}_{ls} - \hat{x}_{pinv} \| = \| A^{+} e \|_2^2 = \| V \Sigma^{-1} U^T e \|_2^2$$

Suppose $\|e\|_2^2 = 1$, what is the worst case reconstruction error?

$$\max_{\substack{e \in \mathbb{R}^m \\ \|e\|_2 = 1}} \| V \Sigma^{-1} U^T e \|_2^2$$

Since the columns of U are orthonormal, $\|U^T e\|_2^2 \le \|e\|_2^2$, and we can find

$$\max_{\substack{\beta \in \mathbb{R}^R \\ \|\beta\|_2 = 1}} \| V \Sigma^{-1} \beta \|_2^2$$

$$= \langle V\Sigma^{-1}\beta, V\Sigma^{-1}\beta \rangle = \langle \Sigma^{-1}\beta, \cancel{V^T V}^{I} \Sigma^{-1}\beta \rangle$$

$$= \| \Sigma^{-1} \beta \|_2^2$$

$$= \max_{\substack{\beta \in \mathbb{R}^R \\ \|\beta\|_2 = 1}} \| \Sigma^{-1} \beta \|_2^2$$

The max is achieved when β has a "1" in the position corresponding to the largest value in Σ^{-1}

$$\underline{\beta}_{max} = \begin{cases} \beta_R = 1 \\ \beta_i = 0 \text{ for } i \neq R \end{cases}$$

$$= \frac{1}{\sigma_R^2}$$

If $\|e\|_2 \neq 1$ then we can perform the above proof using $\hat{e} = \frac{e}{\|e\|_2}$ then to find the max for e, substitute $e = \hat{e}\,\|e\|_2$ at the end yielding

$$\boxed{\| \hat{x}_{ls} - \hat{x}_{pinv} \|_2^2 = \| V \Sigma^{-1} U^T e \|_2^2 \le \frac{1}{\sigma_R^2} \|e\|_2^2}$$

The smallest singular value causes the greatest growth in the reconstruction error due to observation error ($y = Ax + \underline{e}$)

Using the same approach we can bound the error from below...

$$\boxed{\frac{1}{\sigma_1^2}\|e\|_2^2 \le \| \hat{x}_{ls} - \hat{x}_{pinv} \|_2^2 \le \frac{1}{\sigma_R^2} \|e\|_2^2}$$

Board Notes cont.

Tuesday, April 4, 2017 8:32 AM

If we assume the error term is additive Gaussian while noise, that is

$$e[m] \sim \text{Normal}(0, \nu^2)$$

and each entry in e is independent then

$$E\left[\|e\|_2^2\right] = M\nu^2$$

see the discussion on slide set 11 - Linear Inverse Problems

and

$$E\left[\|A^{+}e\|_2^2\right] = \nu^2 \cdot \operatorname{trace}\left(A^{+T} A^{+}\right)$$

$$= \nu^2 \sum_{r=1}^{R} \frac{1}{\sigma_r^2}$$

Truncated SVD

Wednesday, March 29, 2017 9:29 AM

so we have $A = U \Sigma V^T$

$$= \begin{bmatrix} \\ U \\ \\ \end{bmatrix} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_R \end{bmatrix}}_{\Sigma} \begin{bmatrix} & V^T & \end{bmatrix}$$

The textbook uses

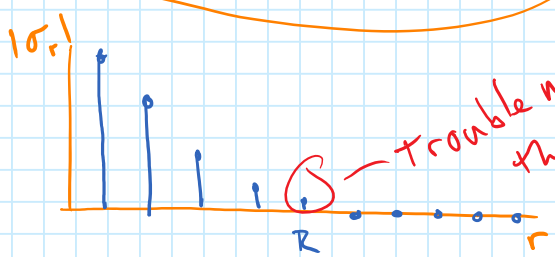
$$\underbrace{\left[\begin{array}{c:c} U & U_0 \end{array}\right]}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_R & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}}_{\Sigma} \underbrace{\left[\begin{array}{c} V^T \\ \hdashline V_0^T \end{array}\right]}_{V^T}$$

selected as discussed before to be orthogonal to U, and orthonormal

σ_1 → largest singular value

smallest singular value → σ_R

U: This is composed of an orthonormal basis for $\mathbb{R}^M$ or $\mathbb{R}^N$ whichever is larger



trouble maker → amplifies the reconstruction error

back to our SVD definition
Σ is $R \times R$

$y = Ax + e \qquad A \in \mathbb{R}^{M \times N}$

$A = U \Sigma V^T \quad , \quad A^{+} = V \Sigma^{-1} U^T$

$$A = \sum_{r=1}^{R} \sigma_r \underline{u}_r \underline{v}_r^T$$

→ A as a sum of rank 1 matrices (similar to the expression of sym + def matrices using eigenvalues and eigenvectors)

$\underline{u}_r \in \mathbb{R}^M$ from U

$\underline{v}_r \in \mathbb{R}^N$ from V

$$A^{+} = \sum_{r=1}^{R} \frac{1}{\sigma_r} \underline{v}_r \underline{u}_r^T$$

order swapped

Error in Truncated SVD

Wednesday, March 29, 2017 9:48 AM

$$\hat{x}_{ls} = A^{\dagger} y = \sum_{r=1}^{R} \frac{1}{\sigma_r} \langle y, u_r \rangle v_r$$

a small $\sigma_r \Rightarrow$ large $\frac{1}{\sigma_r}$ = large error

$$= \sum_{r=1}^{R} \frac{1}{\sigma_r} \langle Ax, u_r \rangle v_r + \sum_{r=1}^{R} \frac{1}{\sigma_r} \langle e, u_r \rangle v_r$$

for small σ_r this is small

so we truncate by choosing $R' < R$

$$\hat{x}_{trunc} = A'^{\dagger} y = \sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle y, u_r \rangle v_r$$

$$\hat{x}_{trunc} - \hat{x}_{pinv} = A'^{\dagger} y - A^{\dagger} A x$$

$$= A'^{\dagger} A x + A'^{\dagger} e - A^{\dagger} A x$$

$$= \boxed{\underbrace{(A'^{\dagger} - A^{\dagger})}_{= \sum_{r=R'+1}^{R} -\frac{1}{\sigma_r} v_r u_r^T} A x} - A'^{\dagger} e$$

$$= \sum_{r=R'+1}^{R} \frac{-1}{\sigma_r} \langle Ax, u_r \rangle v_r$$

$$= \sum_{r=R'+1}^{R} \frac{-1}{\sigma_r} \Big\langle \sum_{j=1}^{R} \underbrace{\sigma_j \langle x, v_j \rangle}_{\text{scalars}} u_j, u_r \Big\rangle v_r$$

$$= \sum_{r=R'+1}^{R} \frac{-1}{\cancel{\sigma_r}} \cancel{\sigma_r} \langle x, v_r \rangle v_r = \sum_{r=R'+1}^{R} - \langle x, v_r \rangle v_r$$

$$A'^{\dagger} e = \sum_{r=1}^{R'} \frac{1}{\sigma_r} \langle e, u_r \rangle v_r$$

Approximation Error

Wednesday, March 29, 2017 10:03 AM

These are mutually orthogonal

$$\hat{x}_{trunc} - \hat{x}_{pinv} = \underbrace{\sum_{k=R'+1}^{R} -\langle x, v_k\rangle v_k}_{\text{approximation error}} + \underbrace{\sum_{r=1}^{R'} \frac{1}{\sigma_r}\langle e, u_r\rangle v_r}_{\text{noise error}}$$

this means that

$$\|\hat{x}_{trunc} - \hat{x}_{pinv}\|_2^2 = \|\text{Approx. error}\|_2^2 + \|\text{noise error}\|_2^2$$

$$= \sum_{k=R'+1}^{R} |\langle x, v_r\rangle|^2 + \sum_{r=1}^{R'} \frac{1}{\sigma_r^2} |\langle e, u_r\rangle|^2$$

← maximized for $r = R'$

we can let $\hat{e} = \frac{e}{\|e\|}$

max error occurs when e is aligned with $u_{R'}$

Assuming, as before, that $e[m] \sim N(0, \nu^2)$

$$E\left[\|\text{Noise error}\|_2^2\right] = \nu^2 \sum_{r=1}^{R'} \frac{1}{\sigma_r^2} \quad \leftarrow \text{expected error}$$

Aside

$$\|x\|_2^2 = \langle x, x\rangle$$

$$\sum_{r=1}^{R} \underbrace{\frac{1}{\sigma_r}\langle e, u_r\rangle}_{\text{scalar}} v_r$$

— the error due to noise $\hat{x}_{trunc} - \hat{x}_{pinv}$

$$\gamma_r = \frac{1}{\sigma_r}\langle e, u_r\rangle$$

$$\sum_r \gamma_r \underline{v}_r \Rightarrow \|\text{error}\|_2^2 = \left\langle \sum_r \gamma_r \underline{v}_r, \sum_k \gamma_k \underline{v}_k \right\rangle$$

$$= \sum_r \gamma_r \left\langle v_r, \sum_k \gamma_k \underline{v}_k \right\rangle$$

$$\sum_r \gamma_r \sum_k \gamma_k v_k^T v_r$$

$$\langle v_r, v_k\rangle = \begin{cases} 1 & r = k \\ 0 & \text{else} \end{cases}$$

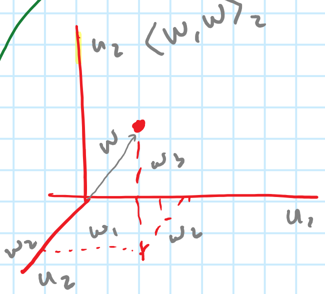
$$= \sum_r \gamma_r^2 = \sum_r \frac{1}{\sigma_r^2}\langle e, u_r\rangle^2$$

if $e = u_R$

$$= \frac{1}{\sigma_R^2}\langle u_R, u_R\rangle^2$$

$\frac{\|u_R\|_2^2}{\sigma_R} \cdot \|u_R\|^2$

normalization factor



$\langle w, w\rangle_2^2 = w_1^2 + w_2^2 + w_3^2$

$\langle w, u_i\rangle = w_i$

$\sum \langle w, u_i\rangle^2 = w_1^2 + w_2^2 + w_3^2$

Aside - Parseval again

Monday, April 3, 2017 9:07 AM

– from a few weeks back

$$\|A^{-1}e\|_2^2 = \|V\Lambda^{-1}V^Te\|_2^2$$
$$= \langle V\Lambda^{-1}V^Te, V\Lambda^{-1}V^Te\rangle$$
$$= \langle \Lambda^{-1}V^Te, \underbrace{\cancel{V^TV}}_{I}\Lambda^{-1}V^Te\rangle$$
$$= \langle \Lambda^{-1}V^Te, \Lambda^{-1}V^Te\rangle \quad (1)$$
$$= \sum_{n=1}^{N} \left|\frac{1}{\lambda_n}\langle e, v_n\rangle\right|^2$$

So viewing the error analysis as a Parseval problem we have

(1)

$$\|\Lambda^{-1}V^Te\|_2^2 = \sum_k |\underbrace{\langle\Lambda^{-1}V^Te, v_k\rangle}|^2$$

$$\to \sum |v_k^T\Lambda^{-1}V^Te|^2$$

$$\sum e^TV\Lambda^{-1}v_kv_k^T\Lambda^{-1}V^Te$$

$$= e^TV\left(\sum_k \Lambda^{-1}v_kv_k^T\right)\Lambda^{-1}V^Te$$

$$= \sum_k \frac{1}{\lambda_k^2}|\langle e, v_k\rangle|^2$$

from 06-Orthogonal Bases

for any orthobasis expansion.

Let S be a Hilbert space with $\langle\cdot,\cdot\rangle_S$ which induces a norm $\|\cdot\|_S$. Let $\{\underline{v}_k\}_{k\in\Gamma}$ be an orthobasis for S. Then for every $\underline{x}, \underline{y} \in S$,

$$\boxed{\langle\underline{x},\underline{y}\rangle_S = \sum_{k\in\Gamma}\alpha_k\overline{\beta}_k}$$

where $\alpha_k = \langle\underline{x}, \underline{v}_k\rangle_S$ and $\beta_k = \langle\underline{y}, \underline{v}_k\rangle_S$

You can think of the α's and β's as transform coefficients of $\underline{x}$ & $\underline{y}$ respectively.

so
$$\langle\underline{x},\underline{y}\rangle_S = \langle\underline{\alpha},\underline{\beta}\rangle_{\ell_2}$$
$$\|\underline{x}\|_S^2 = \|\underline{\alpha}\|_2^2$$

$\sum_k |\langle x, v_k\rangle|^2$

notice, no $\frac{1}{2\pi}$ factor

Thus, every Hilbert space with an orthobasis is equivalent to ℓ_2 — lengths, angles, relations all map to ℓ_2

A symtdef matrix can be expressed as the sum of weighted outer products as $A = \sum_k \lambda_k v_k v_k^T$

$$\Lambda^{-1}v_kv_k^T\Lambda^{-1}$$

$$\sum \begin{bmatrix}\lambda_1^{-1} & & \\ & \lambda_2^{-1} & \\ & & \ddots\end{bmatrix}\begin{bmatrix} | & & \\ v_k & - & v_k^T - \\ | & & \end{bmatrix} = \sum \lambda_k^{-1}v_kv_k^T$$

Notes for Gavin

Monday, April 3, 2017 10:06 AM

When maximizing or minimizing error quantities, we need some constraints. We are really interested in the "direction" of the vector, not its absolute magnitude since we may not know that when performing our analysis of the limits. Therefore, we constrain or normalize $\|e\| = 1$ so we can investigate the impact of the singular values...

Thus, for $\|e\| = 1$, the maximum value for $\|A^+ e\|_2^2$ is $\frac{1}{\sigma_R^2}$. Once we know that, we can find

$$\|A^+ e\|_2^2 \leq \left(\frac{1}{\sigma_R^2}\right) \|e\|_2^2$$

for an unconstrained e

Tikhonov Regularization

Monday, April 3, 2017 10:14 AM

motivation

↪ given A with eigenvalues $\lambda_1, \dots, \lambda_N$

what are the eigenvalues of $(A + \delta I)$?

$$(A + \delta I) v_k = \lambda_k v_k + \delta v_k = (\lambda_k + \delta) v_k$$

regularization

$$\min_{x \in \mathbb{R}^N} \left(\|y - Ax\|_2^2 + \delta \|x\|_2^2 \right)$$

└ penalizes the size of x

- if δ is large, we place great importance on keeping x small.
- if δ is small, we place great importance on finding the best solution to $y = Ax$

How to choose δ?

$$x = V\alpha + V_0 \alpha_0 \qquad y = U\beta + U_0 \beta_0$$

← as in SVD notes

for any x

$$y - Ax = U\beta + U_0\beta_0 - U\Sigma V^T (V\alpha + V_0 \alpha_0)$$

$$= U(\beta - \Sigma\alpha) + U_0 \beta_0$$

since $U_0^T U_0 = I$
$U^T U_0 = 0$
$U^T U = I$

$$\|y - Ax\|_2^2 = \|\beta - \Sigma\alpha\|_2^2 + \|\beta_0\|_2^2$$

and $$\|x\|_2^2 = \|\alpha\|_2^2 + \|\alpha_0\|_2^2$$

$$\|y - Ax\|_2^2 + \delta\|x\|_2^2 = \|\beta - \Sigma\alpha\|_2^2 + \|\beta_0\|_2^2 + \delta\|\alpha\|_2^2 + \delta\|\alpha_0\|_2^2$$

To minimize this for standard least squares, we want $\alpha_0 = 0$
yielding $\beta_0 = 0$

$$\|\beta - \Sigma\alpha\|_2^2 + \delta\|\alpha\|_2^2$$

$$\sum_{k=1}^{R}\left[\left(\beta[k] - \sigma_k\alpha[k]\right)^2 + \delta\,\alpha^2[k]\right] \quad (1)$$

Take the derivative w.r.t. $\alpha[k]$ to minimize ...

$$\frac{\partial}{\partial\alpha[k]}(1) = -2\beta[k]\sigma_k + 2\sigma_k^2\alpha[k] + 2\delta\alpha[k] = 0$$

$$\alpha[k] = \frac{\sigma_k}{\sigma_k^2+\delta}\,\beta[k]$$

$$\hat{\underline{\alpha}}_{tik} = (\Sigma^2 + \delta I)^{-1}\Sigma\beta$$

$$\hat{x}_{tik} = V\hat{\alpha}_{tik} = V(\Sigma^2+\delta I)^{-1}\Sigma U^T y$$

$$\hat{x}_{ls} = V\Sigma^{-1}U^T y = \sum_{r=1}^{R}\frac{1}{\sigma_r}\langle y, u_r\rangle v_r$$

$$\hat{x}_{tik} = \sum_{r=1}^{R}\frac{\sigma_r}{\sigma_r^2+\delta}\langle y, u_r\rangle v_r$$

if $\delta = 0$, these are the same

if $\delta \ll \sigma_r$, $\frac{\sigma_r}{\sigma_r^2+\delta} \approx \frac{1}{\sigma_r}$

if $\sigma \gg \delta_r$, $\frac{\sigma_r}{\sigma_r^2+\delta} \approx 0$

$$\boxed{\hat{x}_{tik} = (A^TA + \delta I)^{-1}A^T y}$$