Bias-Variance Tradeoff in Low-Rank Approximations

Objectives

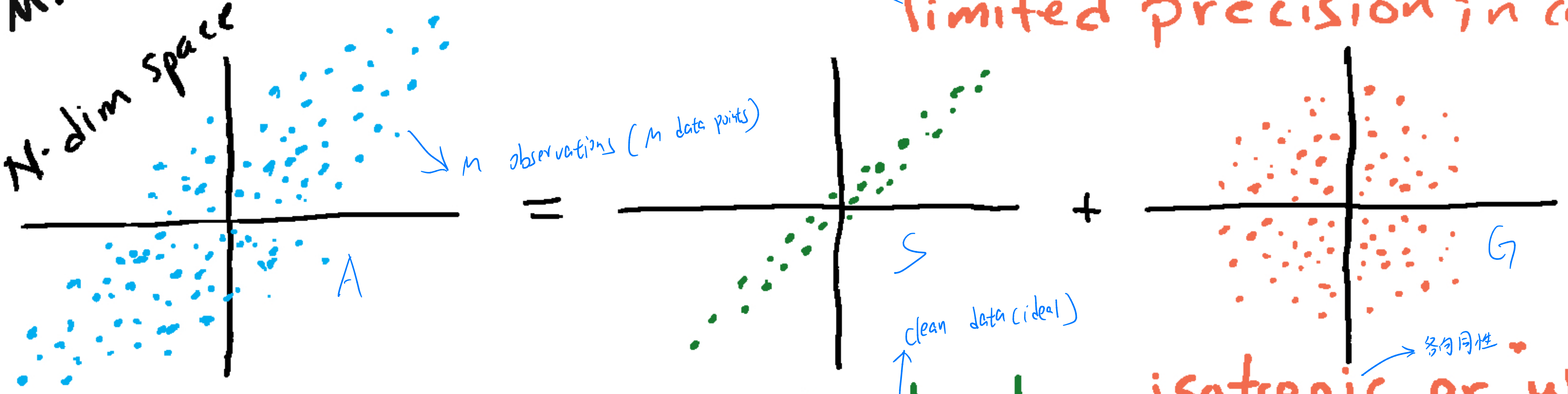
- Introduce concept of noisy data
- Consider impact of noise on SVD
- Define bias and variance
- Use low-rank models to trade bias for variance

Data is often contaminated by noise

$N \times M$, $M > N$

$\underset{\text{measured}}{\underline{A}} = \underset{\text{clean}}{\underline{S}} + \underset{\text{noise}}{\underline{G}}$

- electronics in sensing systems
- enviromental static
- limited precision in computers



diffuse structure

very structured

isotropic or "white" 各向同性

noise – no preferred direction

Sum of squared errors: $\|\underline{G}\|_F^2$

(sum of the squares of all elements in G)

F: the extent

$$\|G\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{M} g_{ij}^2$$

噪声的能量

isotropic

$$\|\underline{G}\|_F^2 = \sum_{i=1}^{N} M\left(\frac{1}{M}\sum_{j=1}^{M} g_{ij}^2\right) = M\sum_{i=1}^{N} \mathrm{var}_i \approx MN\sigma_g^2$$

σ_g^2: variance

var_i is the variance on the dimension i, isomorphic

→ use a standard variance σ_g^2

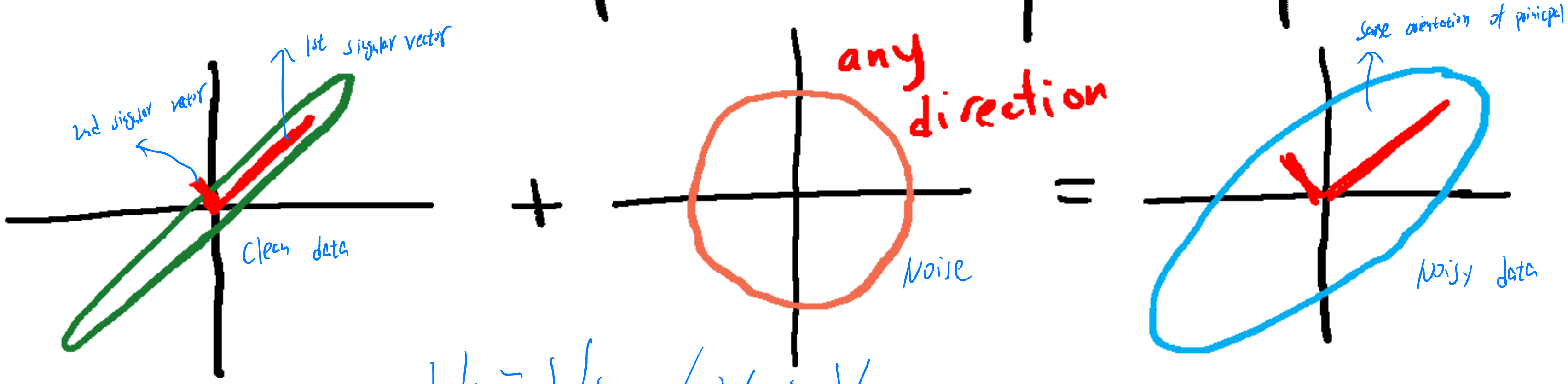
∴ The variance of all dimensions is $N \cdot \sigma_g^2$

∵ we have M data points

∴ $\|G\|_F^2 = MN\sigma_g^2$

Singular vectors are invariant (approx) to isotropic noise

- Proof uses probability concepts



principal comp directions are unchanged ($M \to \infty$)

$U_A \approx U_S$ / $V_A \approx V_S$

variance along each component (singvals) changes

$$\underline{U}\,\underline{\Sigma}_A \underline{V}^T \approx \underline{U}\,\underline{\Sigma}_S \underline{V}^T + \underline{U}\,\underline{\Sigma}_G \underline{V}^T$$

$$\sigma_{A_i} \approx \sigma_{S_i} + M^{1/2}\sigma_g$$

$$\underline{\Sigma}_A = \begin{bmatrix} \sigma_{A_1} & & & 0 \\ & \sigma_{A_2} & & \\ & & \ddots & \\ 0 & & & \sigma_{A_N} \end{bmatrix}, \quad \underline{\Sigma}_S = \begin{bmatrix} \sigma_{S_1} & & & 0 \\ & \sigma_{S_2} & & \\ & & \ddots & \\ 0 & & & \sigma_{S_N} \end{bmatrix}, \quad \Sigma_G \approx \begin{bmatrix} M^{1/2}\sigma_g & & & 0 \\ & M^{1/2}\sigma_g & & \\ & & \ddots & \\ 0 & & & M^{1/2}\sigma_g \end{bmatrix}$$

(isotropic)

$\sigma_{g_i} = M^{1/2} \cdot \mathrm{RMS}$

Low-rank models trade bias for variance

Original: $\underline{A} = \underline{S} + \underline{G}$ Error: $\underline{A} - \underline{S}$, $\|\underline{G}\|_F^2 \approx NM\sigma_g^2$

Low rank: $\hat{\underline{A}}_r = \sum_{i=1}^{r} \sigma_{A_i} \underline{u}_i \underline{v}_i^T \approx \hat{\underline{S}}_r + \hat{\underline{G}}_r$

$$\hat{\underline{S}}_r = \sum_{i=1}^{r} \sigma_{S_i} \underline{u}_i \underline{v}_i^T \qquad \hat{\underline{G}}_r \approx \sum_{i=1}^{r} M^{1/2} \sigma_g \underline{u}_i \underline{v}_i^T$$

Bias^2: $b^2(r) = \|\underline{S} - \hat{\underline{S}}_r\|_F^2$

Error of discarding some information

$$b^2(r) = \Big\| \sum_{i=r+1}^{N} \sigma_{S_i} \underline{u}_i \underline{v}_i^T \Big\|_F^2 = \sum_{i=r+1}^{N} \sigma_{S_i}^2 \quad \text{(notes)}$$

sum of squared "tail" singular values

Variance: $v(r) = \|\hat{\underline{G}}_r\|_F^2$

Error of keeping some bias

$$v(r) = \Big\| \sum_{i=1}^{r} M^{1/2} \sigma_g \underline{u}_i \underline{v}_i^T \Big\|_F^2 = rM\sigma_g^2$$

variance

dimensions × $\frac{\text{variance}}{\text{dimension}}$

Trading bias for variance

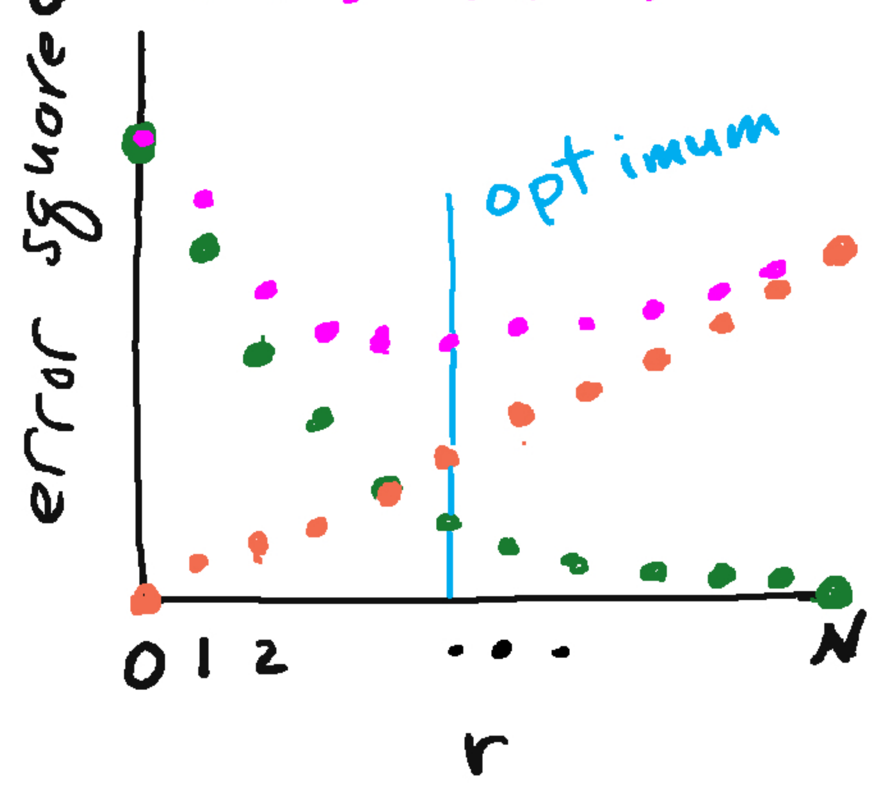
$$b^2(r) = \sum_{i=r+1}^{N} \sigma_{s_i}^2$$ decreases as r increases

丢掉的尾部奇异值的平方和

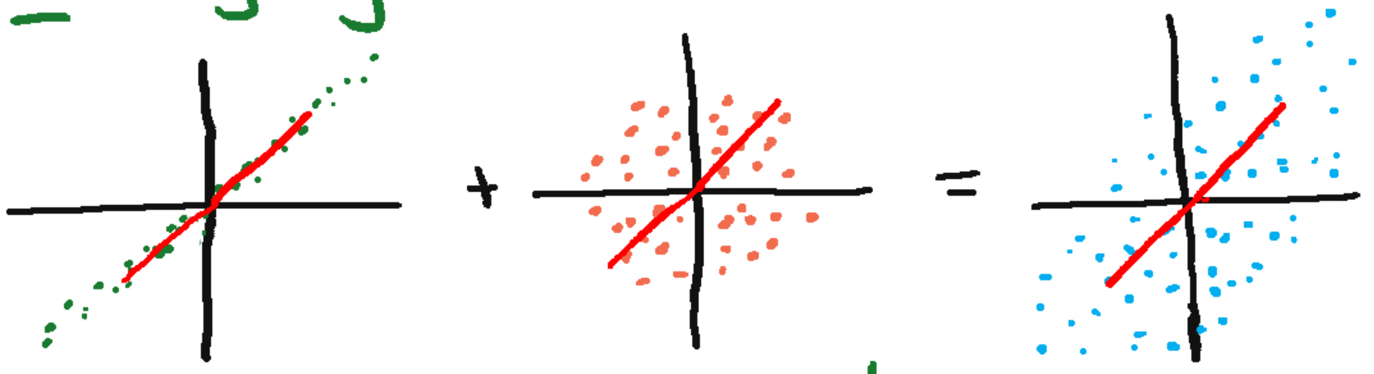
$$v(r) = rM\sigma_g^2$$ increases as r increases

保留下来的r个方向上，噪声成分的大小

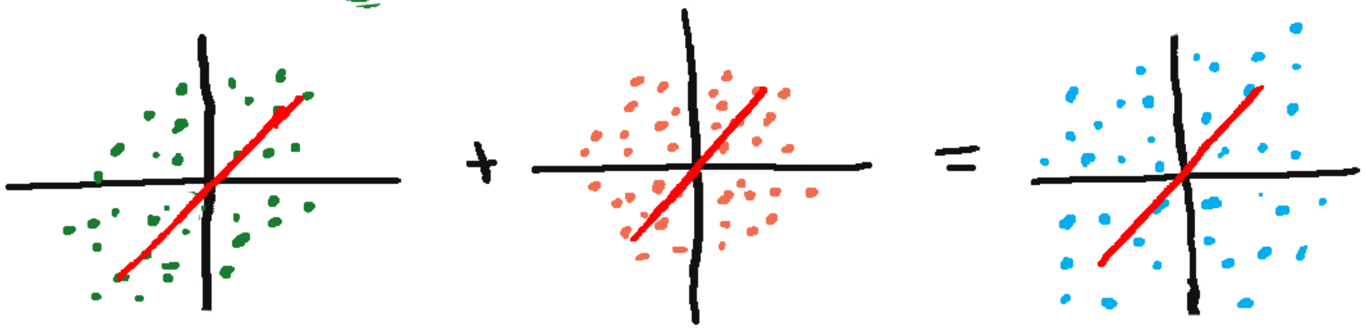
$$b^2(r) + v(r)$$ = Total error



$\underline{S}$ highly structured $\sigma_1 \gg \sigma_2$



$\underline{S}$ weakly structured $\sigma_1 \approx \sigma_2$



Bias-Variance Tradeoff in Low-Rank Representations Proof: Frobenius Norm is the Sum of Squared Singular Values

This proof makes use of the matrix trace operation for square matrices. The trace of a matrix is the sum of the diagonal elements of the matrix. Let $\boldsymbol{B}$ be an M-by-M matrix with elements $B_{i,j}$. Then trace$\{\boldsymbol{B}\} = \text{tr}\{\boldsymbol{B}\} = \sum_{i=1}^{M} B_{i,i}$. One very useful property of the trace operation is that it is invariant to the order of a product of matrices - as long as the products are conformable. Let $\boldsymbol{C}$ be M-by-N with elements $C_{i,j}$ and $\boldsymbol{D}$ be N-by-M with elements $D_{i,j}$. Then both $\boldsymbol{CD}$ and $\boldsymbol{DC}$ are defined. We have

$$\text{tr}\{\boldsymbol{CD}\} = \text{tr}\{\boldsymbol{DC}\}$$

This property follows from the definition of matrix multiplication. The i, i element of $\boldsymbol{CD}$ is $[\boldsymbol{CD}]_{i,i} = \sum_{j=1}^{N} \boldsymbol{C}_{i,j}\boldsymbol{D}_{j,i}$ so

$$\text{tr}\{\boldsymbol{CD}\} = \sum_{i=1}^{M}\sum_{j=1}^{N} \boldsymbol{C}_{i,j}\boldsymbol{D}_{j,i}$$

Similary, the k, k element of $\boldsymbol{DC}$ is $[\boldsymbol{DC}]_{k,k} = \sum_{m=1}^{M} \boldsymbol{D}_{k,m}\boldsymbol{C}_{m,k}$ so

$$\text{tr}\{\boldsymbol{DC}\} = \sum_{k=1}^{N}\sum_{m=1}^{M} \boldsymbol{D}_{k,m}\boldsymbol{C}_{m,k}$$

Interchanging the order of the sums and order of multiplication of the scalars in the sum reveals the equality $\text{tr}\{\boldsymbol{CD}\} = \text{tr}\{\boldsymbol{DC}\}$.

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_M \end{bmatrix}$ be an N-by-M matrix with columns $\boldsymbol{a}_i$. Let the $\boldsymbol{A}$ have singular value decomposition $\boldsymbol{U\Sigma V}^T$ where $\boldsymbol{U}, \boldsymbol{V}$ are square matrices, that is, the full singular value decomposition.

Theorem:

$$||\boldsymbol{A}||_F^2 = \sum_{i=1}^{\min\{M,N\}} \sigma_i^2$$

Proof: First we note that $||\boldsymbol{A}||_F^2 = \sum_{i=1}^{M} \boldsymbol{a}_i^T\boldsymbol{a}_i$ since $\boldsymbol{a}_i^T\boldsymbol{a}_i$ is the sum of the squares of all elements in the i^{th} column of $\boldsymbol{A}$ and we are summing over all M columns.

Next, note that $||\boldsymbol{A}||_F^2 = \sum_{i=1}^{M} \boldsymbol{a}_i^T\boldsymbol{a}_i = \text{tr}\{\boldsymbol{A}^T\boldsymbol{A}\}$ since $\boldsymbol{a}_i^T\boldsymbol{a}_i$ is the i^{th} entry on the diagonal of $\boldsymbol{A}^T\boldsymbol{A}$.

Now substitute the singular value decomposition of $\boldsymbol{A}$ to write

$$||\boldsymbol{A}||_F^2 = \mathrm{tr}\{\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{1}$$

$$= \mathrm{tr}\{\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T\} \tag{2}$$

using the orthonormality of the left singular vectors in $\boldsymbol{U}$. Now move $\boldsymbol{V}$ from the left to the right using the fact that the trace is invariant to the order of a product to obtain

$$||\boldsymbol{A}||_F^2 = \mathrm{tr}\{\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\} \tag{3}$$

$$= \mathrm{tr}\{\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\} \tag{4}$$

$$= \sum_{i=1}^{\min\{M,N\}} \sigma_i^2 \tag{5}$$

where the second line follows from the orthonormality of right singular vectors in $\boldsymbol{V}$ and the last line is a consequence of $\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$ being a square matrix with the squares of the $\min\{M, N\}$ singular values on the diagonal.