SVD and Regularization of Least-Squares Problems

Objectives

- Analyze impact of errors in least-squares problems using SVD
- Introduce truncated SVD regularization
- Analyze ridge regression using SVD

Ill-conditioned least-squares problems have small singular values

$$\min_{\underline{w}} \|\underline{A}\,\underline{w} - \underline{d}\|_2^2 \Rightarrow \underline{w} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d}$$

SVD: $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T \Rightarrow \underline{w} = \underline{V}\,\underline{\Sigma}^{-1}\underline{U}^T\underline{d} = \sum_{i=1}^{P} \frac{1}{\sigma_i}\underline{v}_i(\underline{u}_i^T\underline{d})$

$N \times P$, rank P

$$\Rightarrow \|\underline{w}\|_2^2 = \sum_{i=1}^{P}\left(\frac{1}{\sigma_i}\right)^2(\underline{u}_i^T\underline{d})^2$$

Small $\sigma_i \Rightarrow$ large $\|\underline{w}\|_2$

Prediction with errors: $\tilde{y} = (\tilde{\underline{x}} + \underline{\varepsilon})^T\underline{w} = \underline{x}^T\underline{w} + \underline{\varepsilon}^T\underline{w}$

$$|\underline{\varepsilon}^T\underline{w}|^2 = \|\underline{w}\|_2^2\,\|\underline{\varepsilon}\|_2^2\cos^2\theta$$

large $\|\underline{w}\|_2^2 \Rightarrow$ sensitive to errors

What if rank$(\underline{A}) < P$?

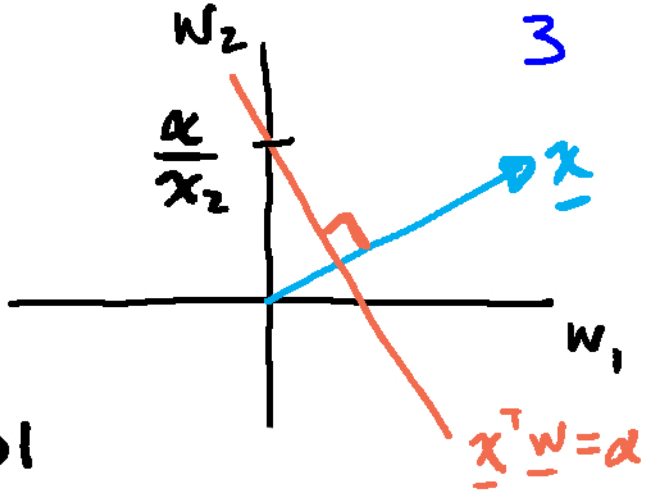
$$\sigma_P = 0$$

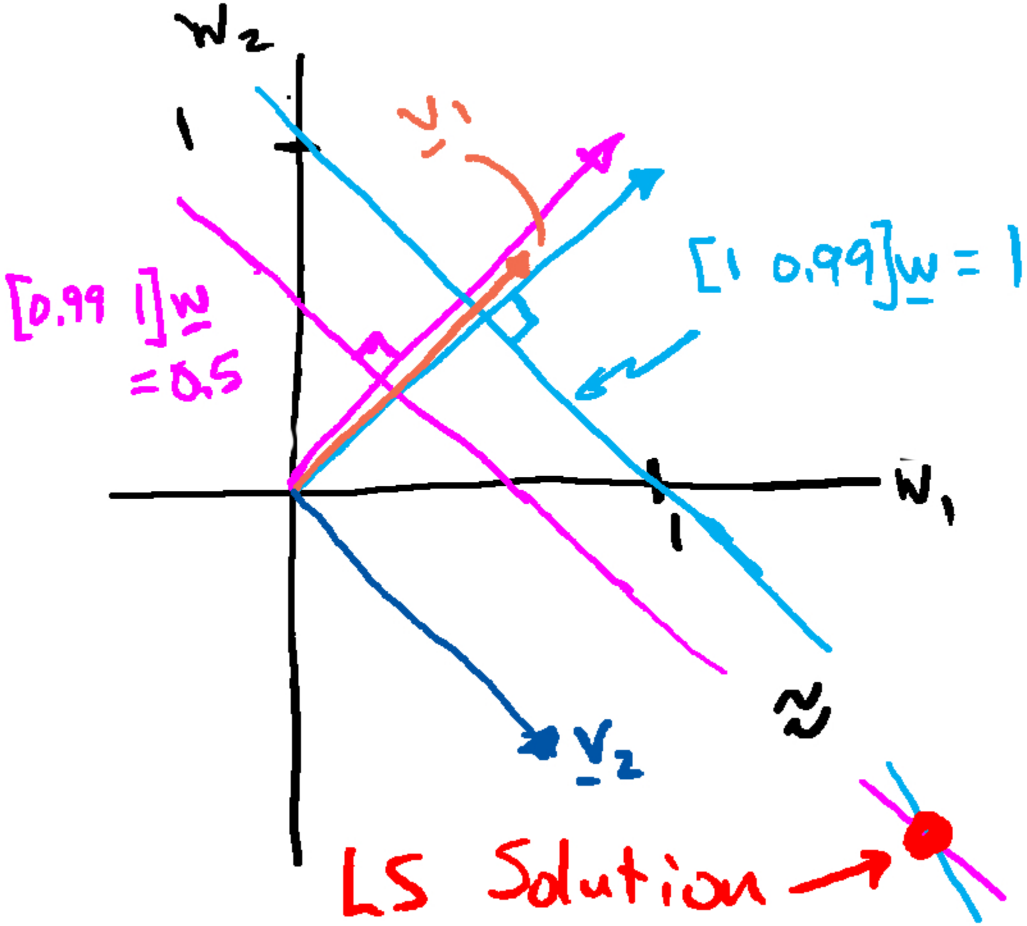
no unique solution

Example: ill-conditioned $\underline{A}$

$$\underline{A} = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}, \quad \underline{d} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

Geometry $\underline{x}^T \underline{w} = \alpha$





$\sigma_1 = 1.99 \quad \sigma_2 = 0.01$

$1/\sigma_1 \simeq 0.5, \quad 1/\sigma_2 = 100$

$\underline{v}_1 = \frac{1}{\sqrt{2}} [1 \quad 1]^T \quad \underline{v}_2 = \frac{1}{\sqrt{2}} [1 \ \ -1]^T$

$\underline{u}_1^T \underline{d} = \frac{1.5}{\sqrt{2}}, \quad \underline{u}_2^T \underline{d} = \frac{0.5}{\sqrt{2}}$

$$\underline{w} = \underline{v}_1 \frac{\underline{u}_1^T \underline{d}}{\sigma_1} + \underline{v}_2 \frac{\underline{u}_2^T \underline{d}}{\sigma_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 3/8 + \begin{bmatrix} 1 \\ -1 \end{bmatrix} 25$$

$\|w\|_2 \uparrow$ as $\sigma_2 \downarrow$ $(\nearrow \rightarrow \nearrow)$

Regularized LS via truncated SVD

Replace $\sum_{i=1}^{p} \frac{1}{\sigma_i} \underline{v}_i (\underline{u}_i^T \underline{d})$ with $\sum_{i=1}^{r} \frac{1}{\sigma_i} \underline{v}_i (\underline{u}_i^T \underline{d})$

where $r < p$.

- Avoid inverting small/zero singular values
- Equivalent to replacing $\underline{A} = \sum_{i=1}^{p} \sigma_i \underline{u}_i \underline{v}_i^T$ with the rank-r approximation $\underline{A}_r = \sum_{i=1}^{r} \sigma_i \underline{u}_i \underline{v}_i^T$
- Increases $\min_{\underline{w}} \|\underline{A}\underline{w} - \underline{d}\|_2^2$
- Can choose r using intuition or cross-validation

Regularized LS via ridge regression

$$\min_{\underline{w}} \|\underline{A}\,\underline{w}-\underline{d}\|_2^2 + \lambda\|\underline{w}\|_2^2 \Rightarrow \underline{w} = (\underline{A}^T\underline{A}+\lambda\underline{I})^{-1}\underline{A}^T\underline{d}$$

controls norm!

Use SVD: $\underline{A}^T\underline{A} = \underline{V}\,\underline{\Sigma}^2\underline{V}^T$, $\lambda\underline{I} = \underline{V}\lambda\underline{I}\,\underline{V}^T$

$$\underline{w} = \left(\underline{V}(\underline{\Sigma}^2+\lambda\underline{I})\underline{V}^T\right)^{-1}\underline{V}\,\underline{\Sigma}\,\underline{U}^T\underline{d} = \underline{V}\underbrace{(\underline{\Sigma}^2+\lambda\underline{I})^{-1}\underline{\Sigma}}_{\underline{D}}\,\underline{U}^T\underline{d}$$

$$\underline{D} = \begin{bmatrix} \frac{1}{\sigma_1^2+\lambda} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_p^2+\lambda} \end{bmatrix}\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{bmatrix} = \begin{bmatrix} \sigma_1/\sigma_1^2+\lambda & & 0 \\ & \ddots & \\ 0 & & \frac{\sigma_p}{\sigma_p^2+\lambda} \end{bmatrix}$$

Controlled!

$$\underline{w} = \sum_{i=1}^{p} \frac{\sigma_i}{\sigma_i^2+\lambda}\,\underline{v}_i(\underline{u}_i^T\underline{d})$$

- as $\sigma_i \to 0$, $\frac{\sigma_i}{\sigma_i^2+\lambda} \to \sigma_i/\lambda$
- increased value $\|\underline{A}\,\underline{w}-\underline{d}\|_2^2$

Copyright 2019
Barry Van Veen