The SVD and Least-Squares Problems

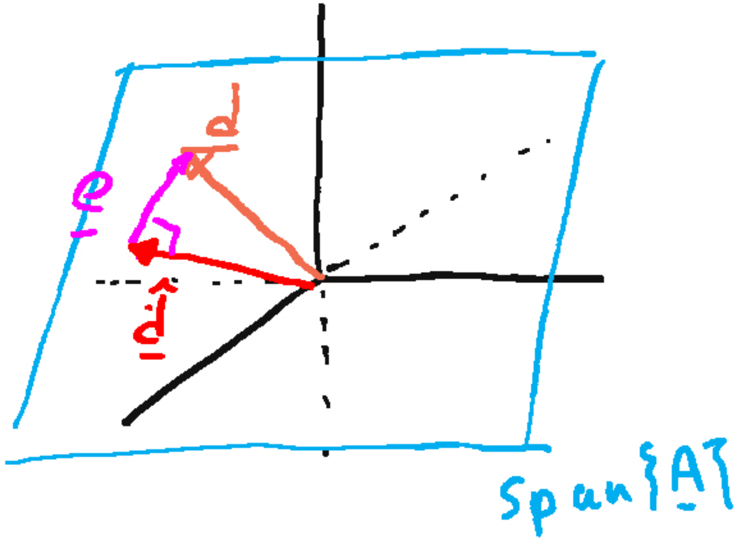
Objectives

- Express least-squares solution in terms of SVD

- Express least-squares error in terms of SVD

- Use SVD to solve the orthobases classification problem

SVD gives insight into the least-squares problem



$$\min_{\underline{w}} \|\underbrace{\underline{d} - \underline{A}\underline{w}}_{\underline{e}}\|_2^2$$

$\underline{A} : N \times p \quad \text{rank}(\underline{A}) = p$

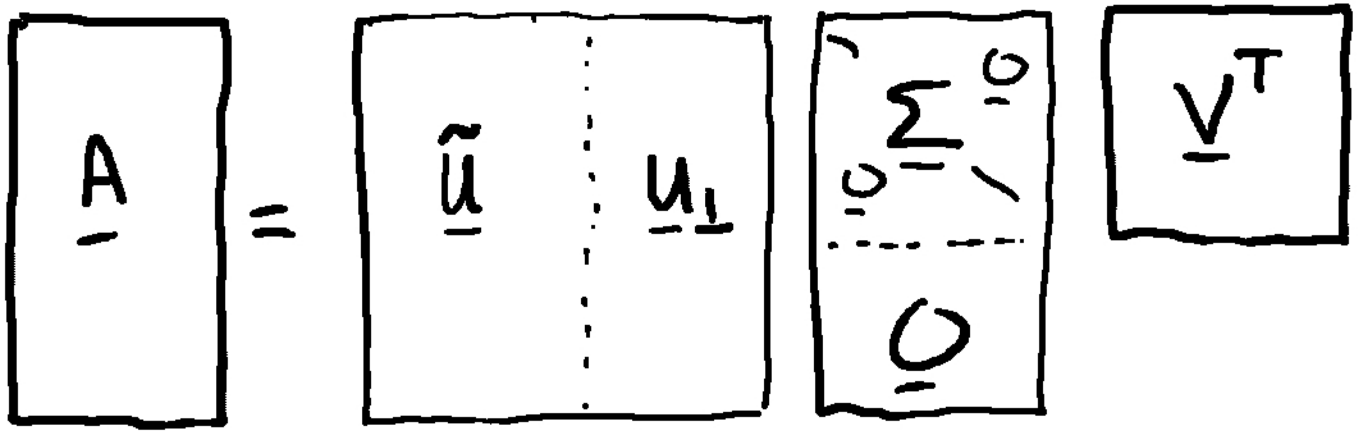
$\underline{d} : N \times 1$

$$\underline{w} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d}$$

$$\hat{\underline{d}} = \underline{A}\underline{w} = \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d} = \underline{P}_A\underline{d}$$

$$\underline{e} = \underline{d} - \hat{\underline{d}} = (\underline{I} - \underline{P}_A)\underline{d} = \underline{P}_{A\perp}\underline{d}$$

SVD -



$$\underline{u} = [\tilde{\underline{u}} \,\vdots\, \underline{u}_\perp] \quad N \times N$$

$$\underline{A} = \tilde{\underline{u}}\,\Sigma\,\underline{V}^T$$

Least-squares solution

$$\underline{w} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d} = (\underline{V}\cancel{\underline{\Sigma}^T}\cancel{\tilde{\underline{u}}^T}\,\cancel{\tilde{\underline{u}}}\,\underline{\Sigma}\,\underline{V}^T)^{-1}\underline{V}\cancel{\underline{\Sigma}^T}\tilde{\underline{u}}^T\underline{d}$$

$$= (\underline{V}\,\underline{\Sigma}^2\underline{V}^T)^{-1}\underline{V}\,\underline{\Sigma}\,\tilde{\underline{u}}^T\underline{d}$$

recall $(\underline{E}\,\underline{F}\,\underline{G})^{-1} = \underline{G}^{-1}\underline{F}^{-1}\underline{E}^{-1}$

$\underline{V}^{-1} = \underline{V}^T$

$$= \underline{V}\,\underline{\Sigma}^{-2}\cancel{\underline{V}^T\underline{V}}\,\underline{\Sigma}\,\tilde{\underline{u}}^T\underline{d}$$

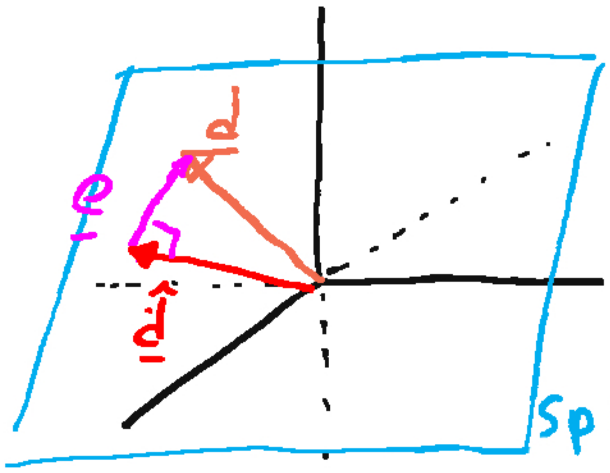
$$= \underbrace{\underline{V}\,\underline{\Sigma}^{-1}\tilde{\underline{u}}^T}_{\text{psuedo-inverse of } \underline{A}}\underline{d} = \sum_{i=1}^{p}\frac{1}{\sigma_i}\underline{v}_i(\tilde{\underline{u}}_i^T\underline{d})$$

$$\underline{\Sigma}^{-2}\underline{\Sigma} = \begin{bmatrix} 1/\sigma_1^2 & & \underline{0} \\ & 1/\sigma_2^2 & \\ & & \ddots \\ \underline{0} & & \frac{1}{\sigma_p^2} \end{bmatrix}\begin{bmatrix} \sigma_1 & & \underline{0} \\ & \sigma_2 & \\ & & \ddots \\ \underline{0} & & \sigma_p \end{bmatrix} = \begin{bmatrix} 1/\sigma_1 & & \underline{0} \\ & 1/\sigma_2 & \\ & & \ddots \\ \underline{0} & & 1/\sigma_p \end{bmatrix} = \underline{\Sigma}^{-1}$$

$$\left[(\underline{A}^T\underline{A})^{-1}\underline{A}^T\right]\underline{A} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{A} = \underline{I}$$

$$\underline{V}\,\underline{\Sigma}^{-1}\tilde{\underline{u}}^T\tilde{\underline{u}}\,\underline{\Sigma}\,\underline{V}^T = \underline{V}\,\underline{\Sigma}^{-1}\underline{\Sigma}\,\underline{V}^T = \underline{V}\,\underline{V}^T = \underline{I}$$

Least-squares error and projections



$$\hat{\underline{d}} = \underline{A}\left[(\underline{A}^T\underline{A})^{-1}\underline{A}^T\right]\underline{d} = \underline{P}_A\,\underline{d}$$

$$= \underline{\tilde{U}}\,\underline{\Sigma}\,\cancel{\underline{V}^T\left[\underline{V}}\,\underline{\Sigma}^{-1}\underline{\tilde{U}}^T\right]\underline{d} = \underline{\tilde{U}}\,\underline{\Sigma}\,\underline{\Sigma}^{-1}\underline{\tilde{U}}^T\underline{d}$$

$$= \underline{\tilde{U}}\,\underline{\tilde{U}}^T\underline{d} \Rightarrow \underline{P}_A = \underline{\tilde{U}}\,\underline{\tilde{U}}^T$$

$$\underline{e} = \underline{d} - \hat{\underline{d}} = (\underline{I} - \underline{\tilde{U}}\,\underline{\tilde{U}}^T)\,\underline{d} = \underline{P}_{A\perp}\underline{d} \Rightarrow \underline{P}_{A\perp} = \underline{I} - \underline{\tilde{U}}\,\underline{\tilde{U}}^T$$

Recall $\underline{U} = [\,\underline{\tilde{U}} \,\vdots\, \underline{U}_\perp\,]$ $(N \times N)$ $\underline{U}^T\underline{U} = \underline{U}\,\underline{U}^T = \underline{I} = [\,\underline{\tilde{U}} \,\vdots\, \underline{U}_\perp\,]\begin{bmatrix} \underline{\tilde{U}}^T \\ \underline{U}_\perp^T \end{bmatrix}$

So $\underline{I} = \underline{\tilde{U}}\,\underline{\tilde{U}}^T + \underline{U}_\perp\underline{U}_\perp^T \Rightarrow \underline{P}_{A\perp} = \underline{U}_\perp\underline{U}_\perp^T$

Classification using SVD o/n bases

Training

$$\begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_N^T \end{bmatrix} \underline{w} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \Rightarrow \underline{A}\,\underline{w} = \underline{d}$$

features   labels

$$\Downarrow$$

Prediction

$$\tilde{y} = \operatorname{sign}(\underline{\tilde{x}}^T \underline{w})$$

orthobases

$$\underline{A} = \underline{\tilde{U}}\,\underline{\Sigma}\,\underline{V}^T$$

$$\underline{\tilde{U}}\,\underline{w}' = \underline{d}$$

$$\Downarrow$$

$$\underline{w}' = \underline{\tilde{U}}^T \underline{d}$$

$$\underline{w} = \underline{V}\,\underline{\Sigma}^{-1}\underline{\tilde{U}}^T\underline{d} = \underline{V}\,\underline{\Sigma}^{-1}\underline{w}'$$

$|\underline{w}_i'|$ indicates importance of i^{th} ortho feature

Orthobases Prediction

$$\tilde{y} = \operatorname{sign}(\underline{\tilde{x}}^T \underline{w}) = \operatorname{sign}(\underline{\tilde{x}}^T \underline{V}\,\underline{\Sigma}^{-1}\,\underline{w}')$$

$$\underline{\tilde{x}}'^T = \underline{\tilde{x}}^T \underline{V}\,\underline{\Sigma}^{-1} \Rightarrow \tilde{y} = \operatorname{sign}(\underline{\tilde{x}}'^T \underline{w}')$$

transformed feature   orthobasis classifier