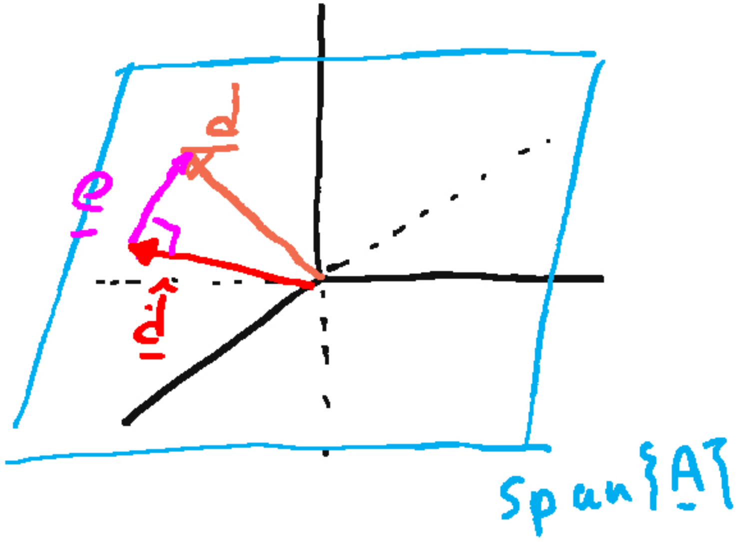
The SVD and Least-Squares Problems

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 -

$$\underset{N\times p}{\underline{A}} = \left[\underset{N\times P}{\tilde{\underline{u}}} \;\vdots\; \underset{N\times(N-P)}{\underline{u}_\perp}\right] \begin{bmatrix} \underline{\Sigma} \\ \underline{0} \end{bmatrix} \underline{V}^T$$

(Goes away when multiplying the zeros in Σ)

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

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

Least-squares solution

Diagonal $\rightarrow \therefore \Sigma^T = \Sigma$ orthonormal, so it is Identity $\underline{A}^T$

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

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

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

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

Because V is orthonormal square matrix

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

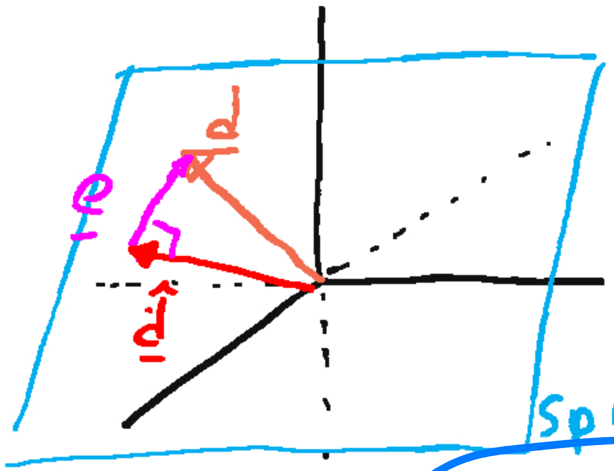
$$\underline{\Sigma}^{-2}\underline{\Sigma} = \begin{bmatrix} 1/\sigma_1^2 & & & 0 \\ & 1/\sigma_2^2 & & \\ & & \ddots & \\ 0 & & & 1/\sigma_p^2 \end{bmatrix}\begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_p \end{bmatrix} = \begin{bmatrix} 1/\sigma_1 & & & 0 \\ &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

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

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

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

Least-squares error and projections



span{A} = span{$\tilde{U}$}

Projection of d onto subspace of A

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

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

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

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

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

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

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

ortho bases

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

Here $w' = \Sigma V^T w$

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

$\Downarrow$

$$\underline{w}' = \tilde{\underline{U}}^T \underline{d} \qquad \tilde{U}^T = \tilde{U}^{-1}$$

$$\underline{w} = \underbrace{\underline{V}\,\underline{\Sigma}^{-1}\tilde{\underline{U}}^T}_{A^{-1}} \underline{d}$$

$$= \underline{V}\,\underline{\Sigma}^{-1}\underline{w}'$$

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

Ortho bases Prediction

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

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

transformed feature   ortho basis classifier