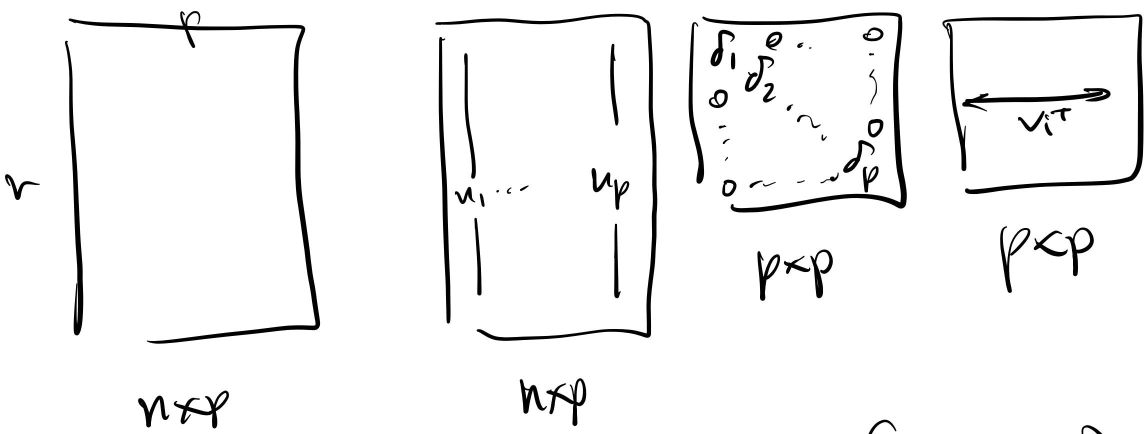
Advanced ML

1/24/23

Suppose we have a matrix $\underset{(n \times p)}{X}$ w/ n rows & p columns ($n > p$). It turns out that *every* matrix X has a *singular value decomposition*:

$$\underset{}{X} = \underset{\text{hanger}}{U} \; \underset{\text{stretcher}}{D} \; \underset{\text{aligner}}{V^T}$$



$V = (v_1 \; v_2 \cdots v_p)$

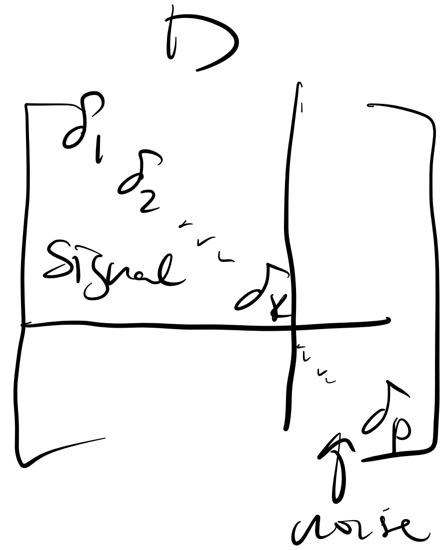
Facts:

1. $U^T U = I_p$
2. $V^T V = I_p$
3. The diagonal entries $\{\delta_1, \ldots, \delta_p\}$ are nonnegative & are called the *singular values* of X.

Alt. Form:

$$X = U D V^T \quad \text{exercise}$$

$$= \sum_{j=1}^{p} \sigma_j u_j v_j^T$$



$$-X = U \, D \, V^T$$

$$= U \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_p \end{pmatrix} V^T$$

$$= U \left(\begin{bmatrix} \sigma_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} + \begin{bmatrix} 0 & & & \\ & \sigma_2 & & \\ & & 0 & \\ & & & \ddots\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \right) V^T$$

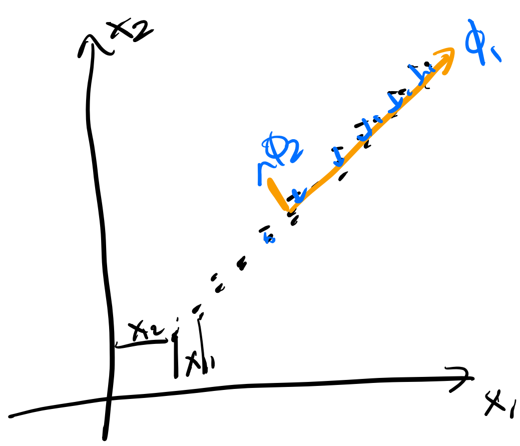
$$= U\begin{bmatrix} \delta_1 & & \\ & 0 & \\ & & \ddots \\ & & & 0 \end{bmatrix} V^T + \cdots + U\begin{bmatrix} 0 & & \\ & \ddots & \\ & & \delta_p \end{bmatrix} V^T$$

$$= \delta_1 u_1 v_1^T + \delta_2 u_2 v_2^T + \cdots + \delta_p u_p v_p^T$$

$$= \sum_{j=1}^{p} \delta_j u_j v_j^T$$

If there are r nonnegative singular values then $\text{rank}(X) = r$.

Recall that if X is a data matrix, then its rank is the dimension of the space in which variation actually exists!



$$D = \begin{pmatrix} 10 & \\ & 0 \end{pmatrix}$$

$$X = \begin{array}{cc} X_1 & X_2 \end{array}$$
$$\begin{pmatrix} X_{11} & X_{12} \\ \vdots & \\ X_{n1} & X_{n2} \end{pmatrix}$$

Practical Question

How do I find out what $U, D, \& V$ are?

$$\underset{n\times p}{X}\ \underset{p\times n}{X^T} = \underset{n\times n}{XX^T}$$

$$
\begin{aligned}
XX^T &= (UDV^T)(UDV^T)^T \\
&= UDV^T(V^T)^T D U^T \\
&= UD\cancel{V^TV}DU^T \\
&= UD^2U^T \leftarrow \text{eigendecomposition of } XX^T
\end{aligned}
$$

Check:

$XX^T u_1 = \text{scalar} \cdot u_1$

How to find V:

$$
\begin{aligned}
\underline{X^TX} &= (UDV^T)^T(UDV^T) \\
&= VD\cancel{U^TU}DV^T \\
&= VD^2V^T \leftarrow \text{eigendecomp of } X^TX
\end{aligned}
$$

Gram matrix

What are the eigenvalues of X^Tx?

$V = (v_1 \ v_2 \ \cdots \ v_p)$

Claim: v_j is an eigenvector of X^TX.

$$X^TX v_j = \underline{(\#)}\ \underline{v_j}$$

$$X^TX v_j = (VD^2V^T) v_j$$

$$= VD^2 \underbrace{\begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_p^T \end{pmatrix} v_j}$$

$$= VD^2 \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$v_j^T v_j = 1$

$v_1^T v_2 = 0$

$$= V \begin{pmatrix} d_1^2 & & & & & \\ & d_2^2 & & & & \\ & & \ddots & & & \\ & & & d_j^2 & & \\ & & & & \ddots & \\ & & & & & d_p^2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= V \begin{pmatrix} 0 \\ \vdots \\ d_j^2 \\ \vdots \\ 0 \end{pmatrix} = d_j^2 v_j$$

Why do we care about SVD?

① SVD can give us insight into other ML methods.

② SVD can help w/ dimension red.
↳ more on PCA later.

Ex:

$$Y = X\beta + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma^2 I)$$

Derive the estimator $\hat{\beta}$ through the SVD of X. You can assume that X has full rank.

Sol:

$$(X^T X)\hat{\beta} = X^T Y$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= \left(\underline{(UDV^T)^T (UDV^T)}\right)^{-1} (UDV^T)^T Y$$

$$= (V D^2 V^T)^{-1} (V D U^T) Y$$

$$= (V^T)^{-1} D^{-2} \cancel{V^{-1} V} D U^T Y$$

$$= V D^{-1} U^T Y$$

How does this relate to multicollinearity & instability in $\hat{\beta}$?

$$\hat{\beta} = V D^{-1} U^T y$$

$$= \sum_{j=1}^{p} d_j^{-1} (u_j^T y) v_j$$

exercise

If d_j very small $\Rightarrow d_j^{-1}$ explodes

$\Rightarrow$ high variance of $\hat{\beta}$.

How does regularization help us solve this instability? (In terms of SVD?)

Reading Assignment: S.3.4 in ESL p64

Exercise: Calculate the L_2-regularized estimate $\hat{\beta}_\lambda$ in terms of the SVD of X.

How does this help stabilize our estimate?