Principal Component Analysis

$$
X = \begin{array}{c}
\begin{array}{ccccccc} \text{age} & \text{height} & \text{weight} & \text{GPA} & \text{income} & \#\text{siblings} & \cdots \end{array} \\
\begin{bmatrix} \vdots & & & & & & \\ & & & & & & \\ & & & & & & \end{bmatrix}_{n \times d}
\end{array}
$$

$n >> d$

PCs : linear combinations of <u>all</u> features in our design matrix

→ capture as much variation as possible

→ minimize projection error

Singular Value Decomposition (SVD)

The SVD, somehow, satisfies these goals for us! It is one way to do PCA (but not the only way).

The SVD can also be used for other things!

X assume is mean centered why? see bottom

$$X = U \Sigma V^T$$

$$X_{n\times d} = \underbrace{U_{n\times d} \Sigma_{d\times d}} V_{d\times d}^T$$

columns of U, V each form an orthonormal set

$$U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{bmatrix} \qquad V = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_d \\ | & | & & | \end{bmatrix}$$

1) $u_i^T u_i = 1$ → orthonormal

$\underbrace{u_i^T u_j = 0 \quad i \neq j}$ ⟶ also apply for V!

$$U^T U = I_d \qquad V^T V = I_d$$

$$X = U \Sigma V^T$$

$$\underbrace{XV}_{\text{principal components}} = \underbrace{U\Sigma}_{\text{principal components}}$$

toy example! not actually possible; cols of V here aren't orthonormal

$$\underbrace{\begin{array}{ccccc} \text{age} & \text{height} & \text{weight} & \text{GPA} & \#\text{siblings} \\ \hline & & & & \\ & & & & \\ & & & & \end{array}}_{X} \qquad \underbrace{\begin{bmatrix} 2 & -3 & \\ 3 & 2 & \\ 1 & 0 & \cdots \\ -4 & 1 & \\ 1 & -2 & \end{bmatrix}}_{V}$$

$$\boxed{PC1 = X v_1 = 2 \cdot \text{age} + 3 \cdot \text{height} + 1 \cdot \text{weight} + (-4) \cdot \text{GPA} + 1 \cdot \#\text{siblings}}$$

$$PC2 = X v_2$$

V contains the "directions" of PCs

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & 0 & \\ & & \sigma_3 & & \\ & 0 & & \ddots \sigma_r \ddots & \\ & & & & 0 \end{bmatrix}_{d \times d}$$

σ_i is not "Standard Deviation i"

σ_i "singular value i"

Singular values of X are the square roots of the eigenvalue of $X^T X$

Variance of PC i

$$= \frac{\sigma_i^2}{n}$$

→ Note : σ_i's are arranged in decreasing order:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$$

- r non-zero singular values means design matrix is of rank r

→ As a result, PCs are arranged from most variance to least variance

Allen's PCA cheat sheet

1) The i^{th} principal component

i^{th} column of XV

i^{th} column of $U\Sigma$

$X v_i$

$\sigma_i u_i$

2) The direction of the i^{th} principal component

i^{th} column of $V = v_i$

3) The rank of X

\# of non-zero singular values

= # non-zero σ_i 's

4) The variance explained by PC i

$$Var(PC\ i) = \frac{\sigma_i^2}{n}$$

→ Note: $Var(PC\ i) \geq Var(PC\ j)$ when $i \geq j$

5) The proportion of variance explained by PC i

$$\frac{Var(PC\ i)}{Var(Total)} = \frac{\frac{\sigma_i^2}{n}}{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n} + \dots + \frac{\sigma_r^2}{n}}$$

$$= \boxed{\frac{\sigma_i^2}{\sum_{j=1}^{r} \sigma_j^2}}$$

Why must we mean center?

Let $X = \begin{bmatrix} | & | \\ x_1 & x_2 \\ | & | \end{bmatrix}$ and

$$V = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}$$
$$\quad\; \underset{v_1}{\shortparallel} \quad \underset{v_2}{\shortparallel}$$

$$PC\ 1 = X v_1 = \alpha_1 x_1 + \alpha_2 x_2$$

when $x_1 = x_2 = 0$,

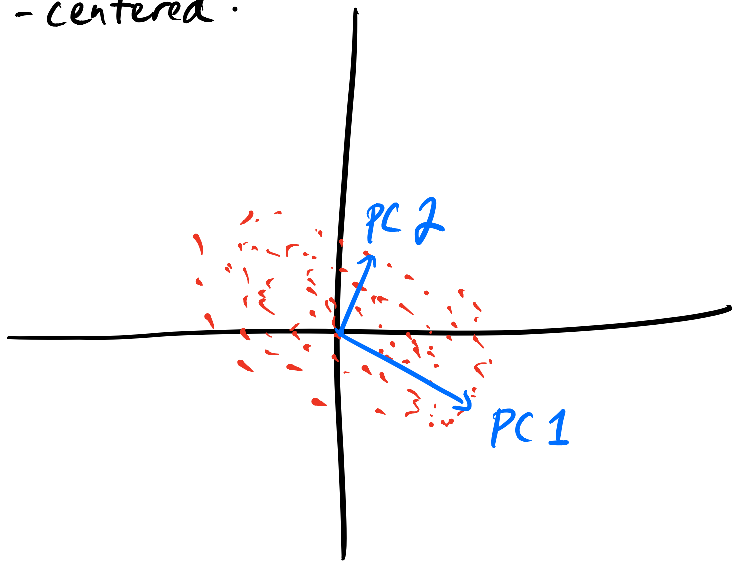
$$\alpha_1 \cdot 0 + \alpha_2 \cdot 0 = 0$$

forcing our PCs to all pass through the origin

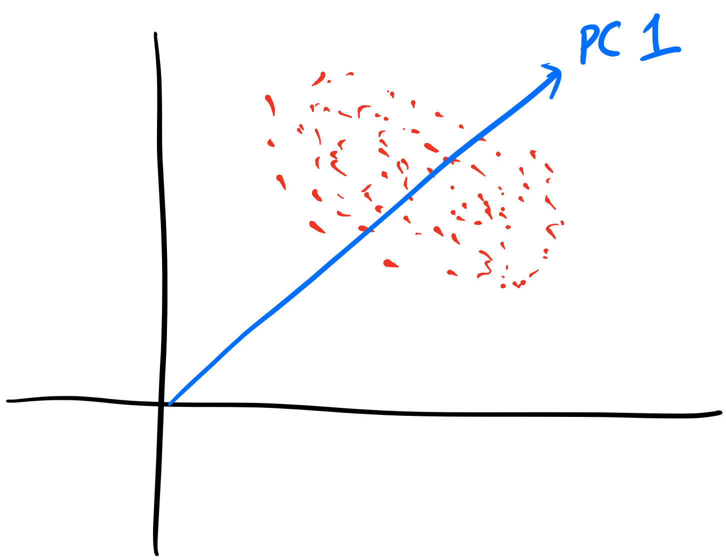
→ Given this constraint, we mean-center to bring our data to the origin.

For example, consider

If mean-centered:



If <u>not</u> mean-centered:



Remember: PCs <u>minimize projection error</u>!