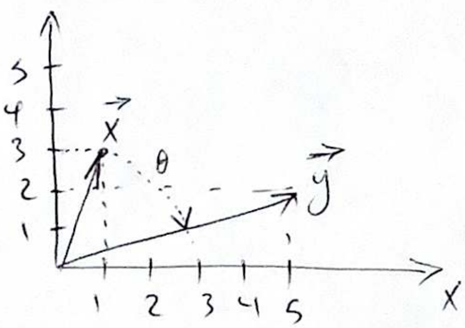
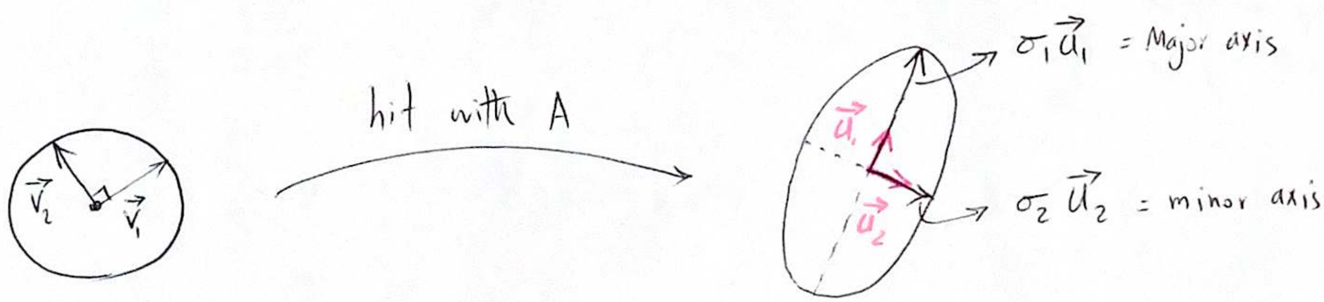
Singular Value Decomposition

$\vec{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ $\quad A = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ $\quad \vec{y} = A\vec{x} = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$



Rotation matrix $= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

scaling matrix $= \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$



Circle

Ellipse

in a more generic way

n-dim $\xrightarrow{\text{hit with some } A \in \mathbb{C}^{m \times n}}$ n-dim

$\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$

$\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n \rightarrow$ Principal axes

$\sigma_1, \sigma_2, \cdots, \sigma_n \rightarrow$ Singular values

$$A\vec{v}_1 = \sigma_1 \vec{u}_1$$

$$A\vec{v}_1 = \sigma_1 \vec{u}_1$$

$$\vdots$$

$$A\vec{v}_j = \sigma_j \vec{u}_j \qquad j = 1, 2, \dots, n$$

$$\underbrace{\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \cdots & \vec{v}_n \end{bmatrix}}_{n \times n} = \underbrace{\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}}_{n \times n}$$

$$AV = \hat{U}\hat{\Sigma}$$

Rotation matrix or unitary transformation

scaling (stretching/compression) matrix

For unitary transformation matrices we have: $V^{-1} = V^*$

$$\hat{U}^{-1} = \hat{U}^*$$

$$(AV = \hat{U}\hat{\Sigma}) \times V^{-1}$$

$$A = \hat{U}\hat{\Sigma}V^* \rightarrow \text{Reduced SVD}$$

After adding silent rows/columns to $\hat{U}$ and $\hat{\Sigma}$ to make them, respectively, $m \times m$ and $m \times n$ we get:

$$\boxed{A = U \Sigma V^*} \rightarrow \text{SVD}$$

When you do SVD, you get the $\underbrace{\sigma\text{'s}}_{\text{singular values}}$ in the Σ always in this order:

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_n \geq 0$$

SVD is guaranteed for every matrix.

Compute U, Σ, and V^*:

$$A^T A = (U \Sigma V^*)^T (U \Sigma V^*)$$
$$= V \Sigma U^* U \Sigma V^*$$
$$= V \Sigma^2 V^*$$

$$A^T A V = V \Sigma^2 V^* V$$

$$A^T A V = V \Sigma^2 \longrightarrow$$ now an eigenvalue problem

Recap: $A\vec{x} = \lambda \vec{x}$

$\det(\lambda I - A) = 0$

once solved $\lambda_j = \sigma_j^2$

$$A A^T = (U \Sigma V^*)(U \Sigma V^*)^T$$
$$= U \Sigma V^* V \Sigma U^*$$
$$(A A^T = U \Sigma^2 U^*) \times U \implies A A^T U = U \Sigma^2$$

This is another eigenvalue problem with the same eigenvalues as the previous case

Principal Component Analysis

Variance

$$\vec{a} = [a_1, a_2, \ldots, a_n] \qquad \vec{b} = [b_1, b_2, \ldots, b_n]$$

$$\sigma_a^2 = \frac{1}{n-1} \vec{a}\,\vec{a}^T$$

$$\sigma_b^2 = \frac{1}{n-1} \vec{b}\,\vec{b}^T$$

Covariance

$$\sigma_{ab}^2 = \frac{1}{n-1} \vec{a}\,\vec{b}^T$$

With X being data matrix:

$$C_X = \frac{1}{n-1} X X^T$$

For example:

$$X = \begin{bmatrix} \bar{x}_a \\ \bar{x}_b \\ \bar{x}_c \\ \bar{x}_d \\ \vdots \\ \bar{x}_f \end{bmatrix}$$

$$C_X = \begin{bmatrix} \sigma^2_{x_a x_a} & \sigma^2_{x_a x_b} & \sigma^2_{x_a x_c} & \cdots \\ \sigma^2_{x_b x_a} & \sigma^2_{x_b x_b} & \sigma^2_{x_b x_c} & \cdots \\ \sigma^2_{x_c x_a} & \sigma^2_{x_c x_b} & \sigma^2_{x_c x_c} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Diagonal entries: measure variance

off-diagonal entries: pair-wise covariance

We'd like to remove redundancy, making C_X diagonal:

$$C_X = \begin{pmatrix} \ddots & & 0 \\ & \ddots & \\ 0 & & \ddots \end{pmatrix} \rightarrow \text{reminds us of SVD}$$

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

Transform data to a new frame of reference using U^*

(Recap: U rotates)

$$Y = U^* X$$

$$C_Y = \frac{1}{n-1} Y Y^T$$

$$= \frac{1}{n-1} U^* X (U^* X)^T$$

$$= U^* X X^T U \left(\frac{1}{n-1}\right)$$

$$= U^* U \Sigma V^* X^T U \left(\frac{1}{n-1}\right)$$

$$= U^* U \Sigma V^* (U \Sigma V^*)^T U \left(\frac{1}{n-1}\right)$$

$$= U^* U \Sigma V^* V \Sigma U^T U \left(\frac{1}{n-1}\right)$$

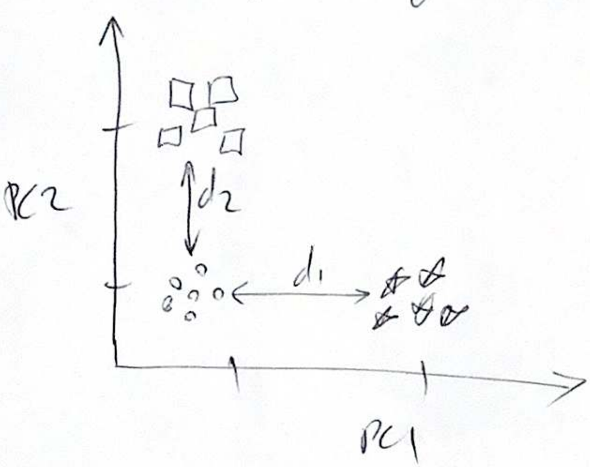
$$C_Y = \frac{1}{n-1} \Sigma^2 \rightarrow \text{now we have a diagonal covariance matrix}$$

$$\lambda_i = \sigma_j^2$$

In PCA, axes are ranked in order of importance

(based on the singular values σ_i)

Example: Imagine 3 clusters



d_1 and d_2 are differences between the clusters.

Differences along the first principal component axis (PC1) are more important than differences along the PC2:

If $d_1 == d_2$, then $\text{cluster}_{\circ}$ and $\text{cluster}_{\star}$ are more different from eachother than $\text{cluster}_{\circ}$ and $\text{cluster}_{\square}$.