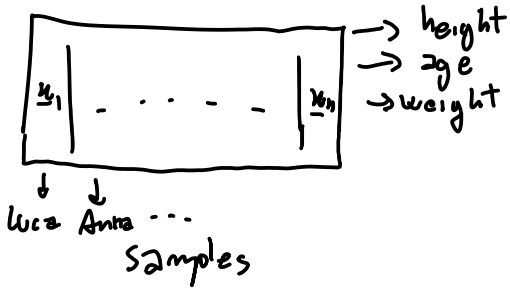
PCA

$X \in \mathbb{R}^{m \times n}$

m # features
n # samples



y_1
y_2
$\vdots$
y_n

Features

- $\mathbb{E}[y_i] \approx \frac{1}{n} \sum_{j=1}^{n} X_{ij} =: \mu_i \qquad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$

$\bar{X}_{ij} := X_{ij} - \mu_i \qquad \bar{X} = X - \mu \, [1 \; \dots \; 1]$

- $Cov(y_h, y_k) := \mathbb{E}[(y_h - \mathbb{E}[y_h])(y_k - \mathbb{E}[y_k])]$

$$\approx \frac{1}{n} \sum_{j=1}^{n} \underbrace{(X_{hj} - \mu_h)}_{\bar{X}_{hj}} \underbrace{(X_{kj} - \mu_k)}_{\bar{X}_{kj}} = \frac{\bar{X}\bar{X}^T}{n}$$

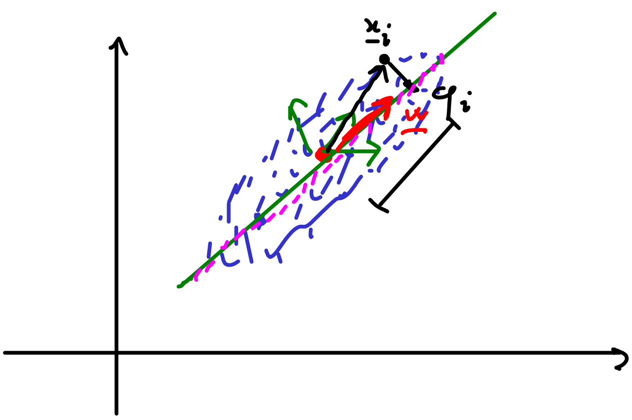
$C = \frac{\bar{X}\bar{X}^T}{n-1} \qquad C_{ii} \approx Var\, y_i \qquad C_{ij} \approx Cov(y_i, y_j)$

$$\boxed{\bar{X} = U \Sigma V^T}$$

total Variance $V := C_{11} + C_{22} + \dots = \sum_{i=1}^{m} C_{ii} = tr(C) = \frac{1}{n-1} \underbrace{tr(\bar{X}\bar{X}^T)}_{\|\bar{X}\|_F^2}$

$\approx Var\, y_1 + \dots + Var\, y_m$

$= \frac{1}{n-1}(\sigma_1^2 + \dots + \sigma_r^2)$



? $w \in \mathbb{R}^m \quad \|w\| = 1$

expaining most of the variance

$\varphi_i := w \cdot (x_i - \mu)$

$\varphi = w^T \bar{X} \in \mathbb{R}^{1 \times m}$

$V_w = \frac{1}{n-1} \varphi \varphi^T$

$* \quad \boxed{\max_{\|w\|=1} V_w}$

$\bar{X} = U \Sigma V^T$

$$V_{\underline{w}} = \frac{1}{n-1}\, \underline{w}^T \bar{X} \bar{X}^T \underline{w} = \frac{1}{n-1}\, \underline{w}^T U \Sigma \underbrace{V^T V}_{\mathbb{1}} \Sigma U^T \underline{w} = \frac{1}{n-1}\, \underline{w}^T U \Sigma^2 U^T \underline{w} \quad *$$

$\underline{w} = U \underline{z}$ $\qquad \|\underline{w}\| = 1 \iff \|\underline{z}\| = 1$

$(\underline{z} := U^T \underline{w})$

$$* \quad \max_{\|\underline{z}\|=1} V(U\underline{z})$$

$$V(U\underline{z}) \overset{*}{=} \frac{1}{n-1}\, \underline{z}^T \underbrace{U^T U}_{\mathbb{1}} \Sigma^2 \underbrace{U^T U}_{\mathbb{1}} \underline{z} = \frac{1}{n-1}\, \underline{z}^T \Sigma^2 \underline{z} = \frac{1}{n-1} \sum_{j=1}^{n} \sigma_j^2 z_j^2$$

$$\left(\sum_{j=1}^{n} z_j^2 = 1 \right)$$

$$\rightarrow \quad \underline{z} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \underline{w} = \begin{bmatrix} \underline{u}_1 & \cdots & \underline{u}_m \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{u}_1$$

$$V_{\underline{u}_1} = \frac{1}{n-1}\, \sigma_1^2 \qquad \qquad \underline{w} = \underline{u}_1, \quad V_{\underline{u}_1} = \frac{1}{n-1}\, \sigma_1^2$$

$(\underline{w}_1, \underline{w}_2)$ $\quad \underline{w}_1 \perp \underline{w}_2$ $\quad$ explaining most of the variance?

$\rightarrow \quad (\underline{u}_1, \underline{u}_2)$

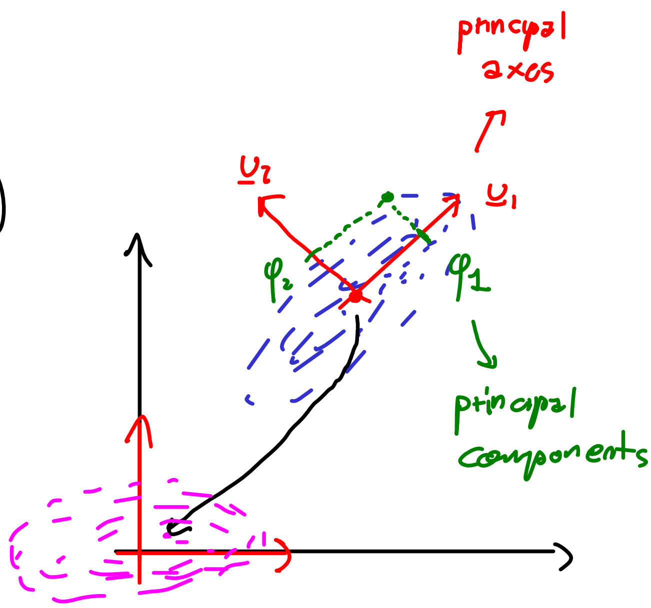
$$V = \frac{1}{n-1} \left(\sigma_1^2 + \sigma_2^2 \right)$$

$$\bar{X} \rightsquigarrow \bar{X}_k := U_k \Sigma_k V_k^T$$

total variance $\quad V_k := \frac{1}{n-1} \left(\sigma_1^2 + \cdots + \sigma_k^2 \right)$

$$\boxed{\frac{V_k}{V} = \frac{\sum_{j=1}^{k} \sigma_j^2}{\sum_{j=1}^{n} \sigma_j^2}}$$

Fraction of "explained" variance



$(\underline{u}_1, \ldots, \underline{u}_r)$

$\Phi_{ij} := \underline{u}_i \cdot (\underline{x}_j - \underline{\mu})$

i-th component of the j-th sample

$$\Phi = U^T \bar{X}$$