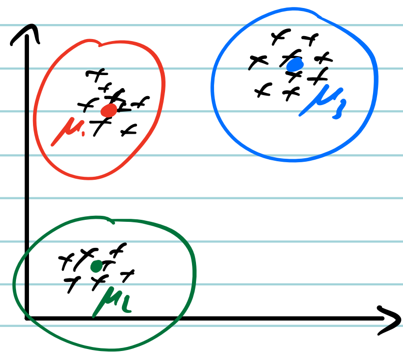
Dimensionality Reduction

Vector Quantisation:



If your data is clustered, you can approximate each input by its cluster assignment. E.g. GMMs give you a probability γ_{ki} that $\vec{x}_i$ is in cluster k.

$\vec{x}_i \rightarrow \begin{pmatrix}\gamma_{1i}\\ \gamma_{2i}\\ \vdots\\ \gamma_{ki}\end{pmatrix} \Leftarrow$ New k-dimensional representation.

Covariances:

Simplifying assumption

Random Variables $X^A, X^B \sim P(X^A, X^B)$ with $\mu^A = E[X^A] = 0$ $\quad \mu^B = E[X^B] = 0$

Variance: $Var(X^A) = E[(X^A - \mu^A)^2] = \underline{E[X^{A2}]}$

Covariance: $Cov(X^A X^B) = E[(X^A - \mu^A)(X^B - \mu^B)] = \underline{E[X^A X^B]}$ $\qquad \underline{COV(X^A, X^A) = VAR(X^A)}$

$$E[X^A X^B] = \begin{cases} > 0 & \text{positively correlated: if } X^A \text{ is } > 0,\ X^B \text{ is } > 0 \text{ (and vice versa)} \\ = 0 & \text{uncorrelated} \\ < 0 & \text{negatively correlated: if } X^A \text{ is } > 0,\ X^B \text{ is } < 0 \text{ (and vice versa)} \end{cases}$$

Covariance Matrix: If $\vec{X} \sim p$ is a vector $\vec{X} = [x_1\ x_2\ x_3\ \dots\ x_d]^T$

Assume data $D = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\} \in \mathbb{R}^d$

$\vec{\mu} = E[X] \approx \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i$ ← weak law of large numbers.

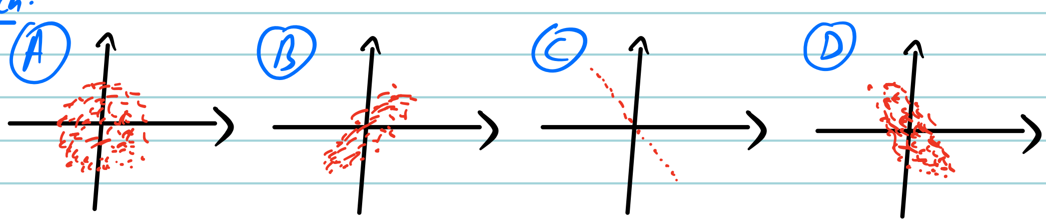
(if $\mu_x = 0$)

$$C^{d\times d} = Cov(\vec{X}) = E[(\vec{X} - \vec{\mu}_x)(\vec{X} - \vec{\mu}_x)^T] = \underline{E[\vec{X}\vec{X}^T]} \approx \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T$$ ← both are indices i.

Covariance Matrix of all r.v. in X, i.e. $X_1, X_2, \dots, X_d$

$C_{\alpha\beta} = COV(x_\alpha, x_\beta)$ $\qquad C_{\alpha\alpha} = VAR[x_\alpha]$
$\alpha \neq \beta$

Match:



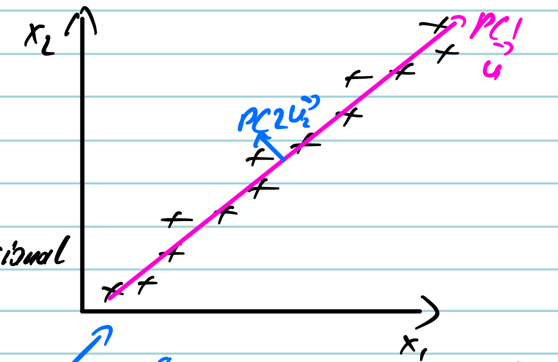
I $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ $\qquad$ II $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

III $\begin{bmatrix} 1 & -0.6 \\ -0.6 & 1 \end{bmatrix}$ $\qquad$ IV $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

Principal Component Analysis:
(Pearson 1901)

Data $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^d$ but are truly from a lower r-dimensional subspace $r \ll d$.

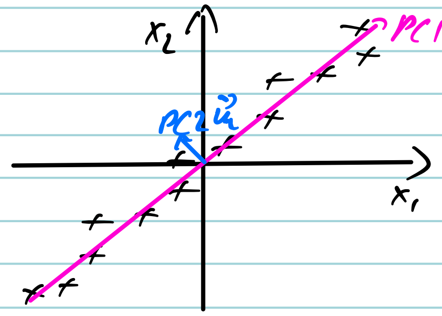
Idea: Find basis vectors for this subspace and project data onto it. $\Rightarrow$ Leads to r-dimensional representation



Everything "interesting" is along PC1
PC2 is likely only noise.

Step#1 of PCA: Center data

$\vec{\mu} = \frac{1}{n}\sum_i \vec{x}_i$ subtract mean: $\vec{x}_i \leftarrow \vec{x}_i - \vec{\mu}$



Step#2: Find first principal component

PCA finds the subspace that contains maximum variance.

PC#1: Find $\vec{v}$ s.t. after projection, $\vec{v}^T\vec{x}_i$, variance is maximized.

$$\max_{\vec{u}^T\vec{u}=1} \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i^T\vec{u})^2 = \max_{\vec{u}^T\vec{u}=1} \sum_i (\vec{x}_i^T\vec{u})^T(\vec{x}_i^T\vec{u}) = \max_{\vec{u}^T\vec{u}} \sum_{i=1}^{n} \vec{u}^T \vec{x}_i \vec{x}_i^T \vec{u} = \max_{\vec{u}^T\vec{u}=1} \vec{u}^T\Big(\underbrace{\sum_{i=1}^{n} \vec{x}_i\vec{x}_i^T}\Big)\vec{u} = \max_{\vec{u}^T\vec{u}=1} \vec{u}^T C \vec{u}$$

we only care about the direction.

Covariance matrix

Lagrangien:

enforces $v^Tv=1$
(if not $\min_\lambda$ will set $\lambda \to \infty$)

$$\max_{\vec{u}} \min_{\lambda} \underbrace{\vec{u}^T C\vec{u} - \lambda(\vec{u}^T\vec{u} - 1)}_{\mathcal{L}(v,\lambda)}$$

$\frac{\partial \mathcal{L}}{\partial v} = 0$ $\Leftarrow$ must hold at optimum.

$\Rightarrow 2C\vec{u} - 2\lambda\vec{u} = 0$

$\therefore \underline{\underline{C\vec{u} = \lambda\vec{u}}}$

u is an eigenvector of C

C has d eigenvectors: $u_1, u_2, \dots, u_d$ s.t. $Cu_i = \lambda_i u_i$
$u_i^T u_i = 1$
$u_i^T u_j = 0$ if $i \neq j$ $\Leftarrow$ orthogonal unit vectors

Sort eigenvectors such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$

u_1 is the first (aka leading) principal component. $\Rightarrow U = [u_1, \dots, u_r]$
u_2 is the second, ...

$$\underline{\text{PCA:}} \quad \underbrace{\vec{x}_i}_{\mathbb{R}^d} \rightarrow \underbrace{U^T(\vec{x}_i - \vec{\mu})}_{\mathbb{R}^r}$$

Reconstruction:

PCA dimensionality reduction: $\vec{z}_i = U^T(\vec{x}_i - \vec{\mu})$

PCA reconstruction: $\hat{x}_i = U\vec{z}_i + \vec{\mu}$

Quiz: proof that if $r=d$ the reconstruction is perfect (i.e. $\hat{x}_i = x_i$).

PCA de-correlates dimensions:

Correlation matrix of $z_1, \dots, z_n$:

$$C_z = \frac{1}{n}\sum_{i=1}^{n} \bar{z}_i \bar{z}_i^T = \frac{1}{n}\sum_i U^T\bar{x}_i (U^T\bar{x}_i)^T = \frac{1}{n}\sum_{i=1}^{n} U^T \bar{x}_i\bar{x}_i^T U = \frac{1}{n}U\underbrace{\left(\sum_{i=1}^{n} \bar{x}_i\bar{x}_i^T\right)}_{C}U$$

$$= \frac{1}{n}U^T C \underbrace{U}_{\text{eigenvectors}} \Rightarrow [C_z]_{ij} = u_i^T C u_j \begin{cases} \lambda_i & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$$

$$C_z = \begin{bmatrix} -u_1^T- \\ -u_i^T- \\ -u_r^T- \end{bmatrix} C \begin{bmatrix} u_1 & u_j & u_r \end{bmatrix} \Rightarrow [C_z]_{ij} = u_i^T C u_j$$

$$\Rightarrow C_z = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$$

How to pick r?

λ_i is the variance within r^{th} PC.

If you project onto r dimensions you lose $\left(\frac{\sum_{i=1}^{r} \lambda_r}{\sum_{i=1}^{n} \lambda_i}\right)$ fraction of the total variance.

Project out 5% variance as noise.

Denoising: Pick smallest r such that $\frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq 0.95$

Singular Value Decomposition:

(for centered data)

$$X = U S V^T$$

U: principal components

S: eigenvalues of C

$SV^T = U^T X \leftarrow$ projected data