PCA & ICA

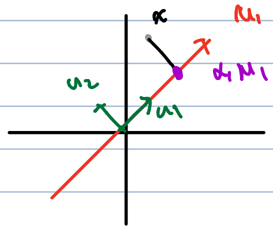
PCA

Preprocess

1. Given $x^{(1)}, \dots x^{(n)} \in \mathbb{R}^d$ $\qquad x^{(i)} \to x^{(i)} - \mu$
2. May need to rescale components (optional)

$$x_j^{(i)} \to \frac{x_j^{(i)} - \mu}{\sigma_j}$$

PCA as optimization



u_1 $\{tu_1, t \in \mathbb{R}\}$: line corresponds to u_1

How to find the closest point to the line?

$$\alpha_1 = \underset{\alpha}{\operatorname{argmin}} \|x - \alpha u_1\|^2$$

$$= \underset{\alpha}{\operatorname{argmin}} \|x\|^2 + \alpha^2 \cancel{\|u_1\|^2}^{1} - 2\alpha(u_1 \cdot x)$$

$$\nabla_{\alpha_1} = 2\alpha - 2(u_1 \cdot x) = 0 \quad \Rightarrow \quad \alpha = u_1 \cdot x$$

Generalize

$u_1, \dots u_k \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$ $\qquad (u_i u_j = \delta_{ij})$

$$\underset{\alpha_1 \dots \alpha_d}{\operatorname{Argmin}} \|x - \sum_{j=1}^{k} \alpha_j u_j\|^2 \Rightarrow \alpha_j = u_j \cdot x$$

$$\|x - \sum \alpha_j u_j\|^2 \leftarrow \text{Residual}$$

We can find PCA by either

1. Maximize Projection subspace
2. Minimize Residuals

$$\max_{\substack{v \in \mathbb{R}^d \\ \|v\| = 1}} \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} \cdot v)^2$$

Solve the optimization problem, we need some facts.

- Let A be symmetric & square

$A = U \Lambda U^T$ in which

- $UU^T = U^T U = I$ (orthonormal basis)
- Λ is diag.

$\Lambda_{ii} = \lambda_i$ and $\lambda_1 > \dots \lambda_n$ (eigenvalues)

Recall If $x = \sum_{i=1}^{d} \alpha_i u_i$ where $[u_1, \dots u_d] = U$

$$Ax = U \Lambda U^T x = U \Lambda \sum_{i=1}^{d} \alpha_i e_i \qquad (u_i u_j = \delta_{ij})$$

$$= U \sum_{i=1}^{d} \alpha_i \lambda_i e_i$$

$$= \sum_{i=1}^{d} \alpha_i \lambda_i u_i$$

fix i, let $c \in \mathbb{R}^+$

$x = c u_i \qquad Ax = \lambda_i x$

$$\max_{\substack{v \in \mathbb{R}^d \\ \|v\| = 1}} \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} \cdot v)^2 = \max_{\substack{U \in \mathbb{R}^d \\ \|U\| = 1}} U^T A U \qquad A = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} x^{(i)T}$$

$$= \max_{\alpha \in \mathbb{R}^d} \sum_{i=1}^{d} \alpha_i^2 \lambda_i$$

What should we pick to maximize?

$\alpha_1 = 1, \quad \alpha_2 = \alpha_3 = \cdots \alpha_d = 0$

- u_1 is the principal eigenvector

what if we want the top-k such vectors?

$u_1 \ldots u_k$ because $\lambda_1 \geq \ldots \geq \lambda_k$

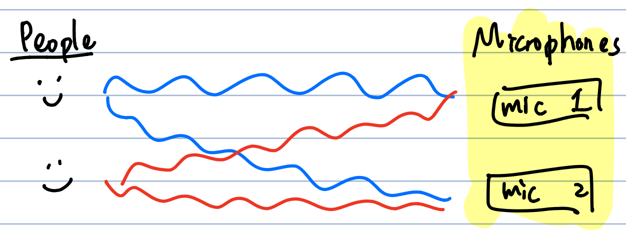
$$x^{(i)} \mapsto \sum_{j=1}^{k} (x^{(i)} \cdot u_j) u_j$$

Notes

I think the CS168 L8 notes and the 229 Class Notes done a better job explaining PCA.

ICA

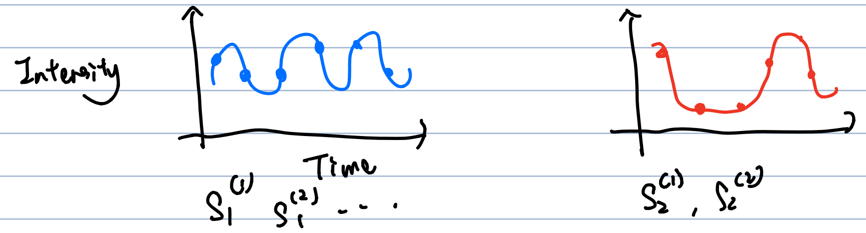
- Indepent Component Analysis



we see a mixture of and

Speaker seperation

Speaker S_1, S_2 Data $x_1^{(t)}, x_2^{(t)}$



$s_j^{(t)}$ is the intensity at time t with speaker j. (NO Observation)

Only observe $x_j^{(t)}$

model $x_j^{(t)} = a_{j1} s_1^{(t)} + a_{j2} s_2^{(t)}$

Microphone j sees a mixture

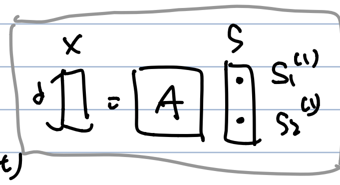
Observed → $x^{(t)} = A s^{(t)}$ ← Latent

For simplicity d is the # of microphones.

Given: $x^{(1)}, x^{(2)}, \dots x^{(n)} \in \mathbb{R}^d$

Do: find $s^{(1)}, \dots s^{(n)} \in \mathbb{R}^d$

And $A \in \mathbb{R}^{d \times d}$ s.t. $x^{(t)} = A \cdot s^{(t)}$



We call A the mixing matrix and $W = A^{-1}$ the unmixing matrix

where $W = \begin{bmatrix} - w_1^T - \\ \vdots \\ - w_d^T - \end{bmatrix}$ So that $s_j^{(t)} = w_j x^{(t)}$

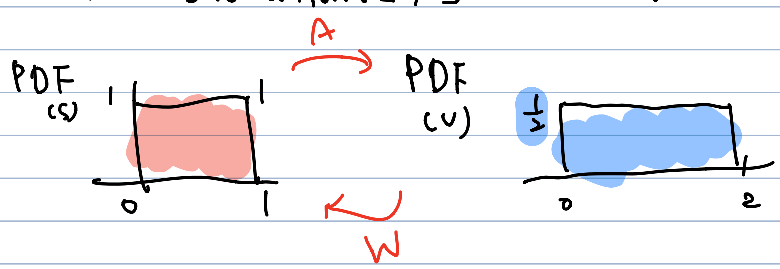
- We center the data $x^{(i)} \to x^{(i)} - \mu$
- A does not vary w/ time, A is full rank
- There is some inherent ambiguity
  - ⇒ we can't determine speaker
  - ⇒ can't determine intensity exactly
- Surprising, speakers cannot be Gaussian

Algorithm

- Just MLE + GD.

Detour Density under linear transform

ex: $S \sim \text{uniform}[0, 1]$ $U = 2S$. whats the PDF of U?



$P_S(x)$
$= \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \end{cases}$

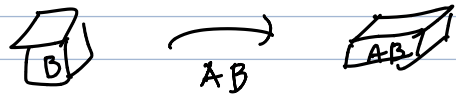
$P_U(x) = P_S(\frac{x}{2}) \cdot \frac{1}{2}$

A : Square and invertible. $U = AS$ $s \sim$ PDF of P_s

$$P_U(x) = P_s(A^{-1}x)\,|\det(A^{-1})|$$

$$= P_s(Wx)\,|\det(W)|$$

- Change of variables formula



Volume B $\mathrm{Vol}(AB) = |\det(A)|\,\mathrm{vol}(B)$

From here ICA is MLE!

latent → $P(S) = \prod_i P_s(S_j)$ "Sources independent and have some distribution"

Observed → $P(X) = \prod_{j=1}^{d} P_s(W_j x)\,|\det(W)|$ (Don't know W)

Key tech!

Set $P(x) \propto g'(x)$ for $g(x) = (1+e^{-x})^{-1}$ (Some magic?)

$$\ell(W) = \sum_{t=1}^{n} \sum_{j=1}^{d} \log g'(w_j x^{(t)}) + \log|\det(W)|$$