m { Finish Factor Analysis
PCA

A little behind our planned pace
But well in buffer!
Please keep questions coming

Factor Analysis Recap

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

Recall: Many fewer points than dimensions "$n \ll d$"

think about: 1000 sensors measured for some ~~small~~ small # of days

Key IDEA: latent structure in r.v. (constrains degree of freedom)

Recall likelihood function

$$\ell(\mu, \Sigma) = \sum_{i=1}^{n} \log \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x^{(i)}-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

$$= \sum_{i=1}^{n} - \frac{1}{2}(x^{(i)}-\mu)^T \Sigma^{-1}(x^{(i)}-\mu) - \log 2\pi |\Sigma|^{1/2}$$

Hence equivalent to look at

$$g(\mu, \Sigma) = \sum_{i=1}^{n} (x-\mu)^T \Sigma^{-1} (x^{(i)}-\mu) + \log|\Sigma| \quad + \cancel{\log 2\pi}$$

We'll reason about structure in Σ

Recall: If $\operatorname{rank}(\Sigma) \le n < d$, many problems! Σ^{-1} is not defined

$|\Sigma| = 0$

Estimate: if Σ is full rank, easy to estimate μ.

$$\nabla_\mu g(\mu, \Sigma) = 2 \sum_{i=1}^{n} \Sigma^{-1} (x^{(i)} - \mu) = 0 \quad \text{(critical points)}$$

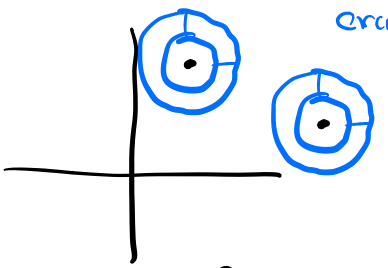
$$\Leftrightarrow \Sigma^{-1} \left(\sum_{i=1}^{n} (x^{(i)} - \mu) \right) = 0 \quad \text{(full rank)}$$

$$\Rightarrow \mu = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} \quad \text{Estimate } \mu$$

$\therefore$ μ is easy to estimate whenever Σ is full rank

Building Block 1 $\quad \Sigma = \sigma^2 I \qquad \sigma^2$ is just a scalar

circular covariance



How many free parameters?

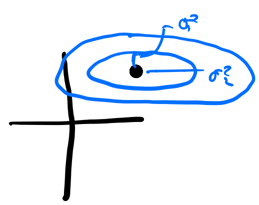
$\Rightarrow$ 1 in Σ (σ^2)

Last time $\quad \sigma^2 = \frac{1}{nd} \sum_{i=1}^{n} \|x^{(i)} - \mu\|^2$

Estimate σ^2 by subtracting mean, squaring entries

Building Block #2

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$



Axis Aligned

$$g(\mu, \Sigma) = \sum_{i=1}^{n} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) + \log |\Sigma|$$

$$|\Sigma| = \prod_{j=1}^{d} \sigma_j^2$$

$$(z_j = \sigma_j^2) = \sum_{i=1}^{n} \sum_{j=1}^{d} z_j^{-1} (x_j^{(i)} - \mu_j)^2 + \log z_j$$

Decouples into d problems.

Fix j $\qquad \sum_{i=1}^{n} z_j^{-1} (x_j^{(i)} - \mu_j)^2 + \log z_j$

$$\sigma_j^2 = \frac{1}{n} \sum_{i=1}^{n} (x_j^{(i)} - \mu_j)^2$$ 1d Gaussians

$\therefore$ if we assume this Axis-Aligned Structure, we can also fit its components

Our Factor Model

Parameters

$$\boxed{\text{Fit}} \quad \left[\begin{array}{l} \mu \in \mathbb{R}^d \\ \Lambda \in \mathbb{R}^{d \times s} \\ \Phi \in \mathbb{R}^{d \times d} \end{array} \right.$$

$s < d$ "small dim"

Φ – Diagonal matrix

Model $P(x, z) = P(x \mid z) P(z)$ z is latent variable

$z \sim N(0, I) \in \mathbb{R}^s$ for $s < d$

$x = \mu + \Lambda z + \epsilon$ or $x \sim N(\mu + \Lambda z, \Phi)$

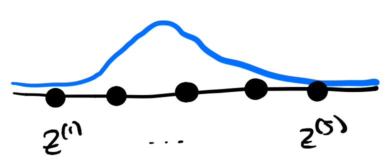
mean (μ)

map from small dim s to high dim (Λz)

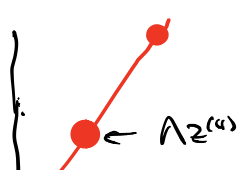
$\epsilon \sim N(0, \Phi)$ noisy.

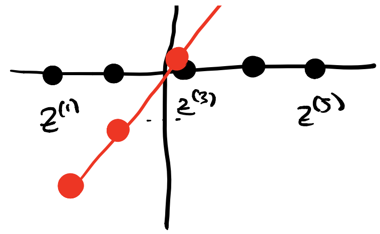
Ex: $d = 2$, $s = 1$, $n = 5$: $x = \mu + \Lambda z + \epsilon$

1. Generate $z^{(1)}, \ldots z^{(5)}$ from $N(0, 1)$



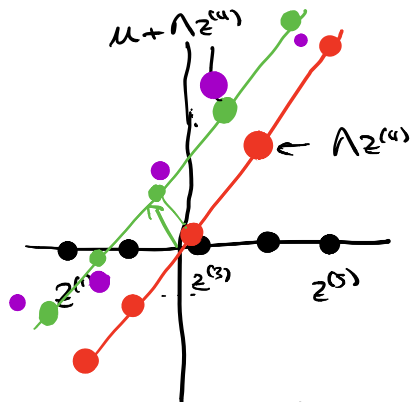
2. $\Lambda = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ Λz





3. Add μ

$x = \mu + \Lambda z + \epsilon$



Finally our data purple which full dimensional

Model is clear: Small dim $\rightarrow$ large dim

TECHNICAL TOOLS Block Gaussians

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad x_1 \in \mathbb{R}^{d_1} \quad x_2 \in \mathbb{R}^{d_2} \quad x \in \mathbb{R}^{d_1+d_2}$$

$$\Sigma = \begin{array}{c} \begin{array}{cc} d_1 & d_2 \end{array} \\ \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \end{array} \begin{array}{l} d_1 \\ d_2 \end{array} \qquad \Sigma_{ij} \in \mathbb{R}^{d_i \times d_j} \qquad i,j \in \{1,2\}$$

Fact 1: $P(x_1) = \int_{x_2} P(x_1, x_2)$ MARGINALIZATION

For Gaussians, $p(x_1) = \mathcal{N}(\mu_1, \Sigma_{11})$ **Gaussian**

Fact 2: $P(x_1 | x_2) \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$ **conditions**

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{11}$$

Matrix inversion lemma

Summary: learned about block notation & we say Gaussians are closed under marginalization & conditioning.

Factor Analysis EM

$$x = \mu + \Lambda z + \epsilon$$

$$\begin{pmatrix} z \\ x \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \Sigma\right)$$

since $\mathbb{E}[z] = 0$

$\mathbb{E}[x] = \mu$

$$\Sigma_{11} = \mathbb{E}[zz^T] = I \qquad (z \sim N(0, I))$$

$$\Sigma_{12} = \mathbb{E}[z(x-\mu)^T] = \mathbb{E}[z z^T \Lambda^T] + \mathbb{E}[\cancel{z\epsilon^T}]^{0}$$

$$= \Lambda^T$$

$$\Sigma_{21} = \Sigma_{12}^T \qquad (\Sigma \text{ is symmetric})$$

$$\Sigma_{22} = \mathbb{E}[(x-\mu)(x-\mu)^T]$$

$$= \mathbb{E}[(\Lambda z + \epsilon)(\Lambda z + \epsilon)^T]$$

$$= \mathbb{E}[\Lambda z z^T \Lambda^T] + \mathbb{E}[\epsilon\epsilon^T]$$

$$= \Lambda\Lambda^T + \Phi$$

Diagonal matrix.

$$\Sigma = \Big[\ I \qquad \Lambda^T$$

$$[\Lambda \quad \Lambda\Lambda^T + \Phi]$$

EM Algorithm?

- **E-step**: $Q_i(z) = P(z^{(i)} \mid x^{(i)}, \theta)$ — use conditional
- **M-step**: We know how to estimate.

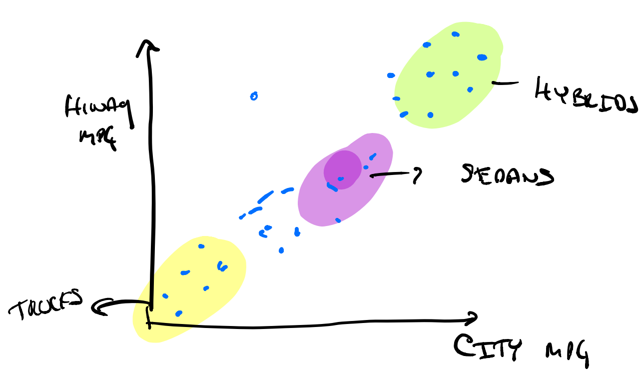
Summary:

- We saw that EM captured GMM
- We learned Factor Analysis (FA)
- We saw that estimates params of FA with EM

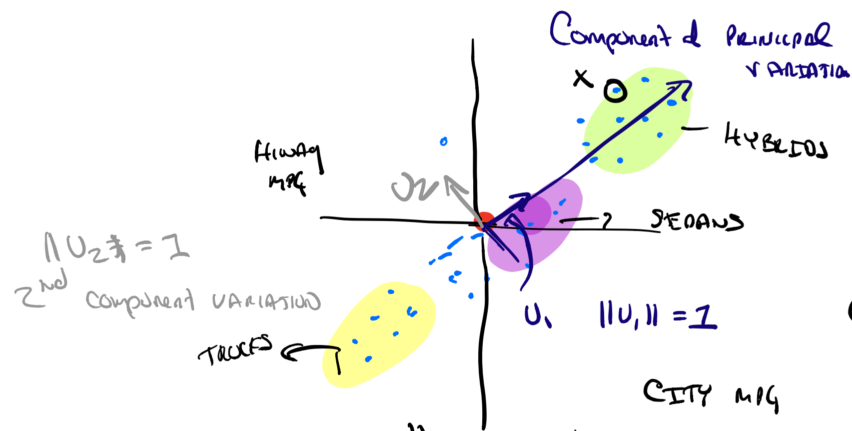
PCA: Principal Component Analysis

Structure	Prob.	Non-Prob.
"Cluster"	GMM Factor	K-means
"Subspace"	Factor Analysis	PCA

Ex: Given pairs about cars (Highway MPG, City MPG)



Question: "Good MPG" $\qquad x^{(1)} \dots x^{(n)} \in \mathbb{R}^d$



CENTER DATA

$$\mu = \frac{1}{N} \sum_i x^{(i)}$$

$$x^{(i)} \mapsto x^{(i)} - \mu$$

$$u_1 \cdot u_2 = 0$$

$$u_i \cdot u_j = \delta_{ij} \quad \text{(orthonormal basis)}$$

u_1 is "How good is MPG"

u_2 is "Difference Hiway and city"

$$x = \alpha_1 u_1 + \alpha_2 u_2$$

$\alpha_1 u_1$ → Keeping this one term

$x^{(i)}$ keep around $\alpha_1^{(i)}$ as one number instead of 2.

What we're think about is when we have 1000s of directions and we want to find a few important ones

$\mathbb{R}^{1000} \to \mathbb{R}^{10}$ or $\mathbb{R}^2$ for visualization.

u_i s.t. $\|u_i\| = 1$ and $u_i \cdot u_j = \delta_{ij}$ $\quad i = 1 \dots n \quad u_j \in \mathbb{R}^n$

$x = \sum_i \alpha_i u_i$ (write in a basis)