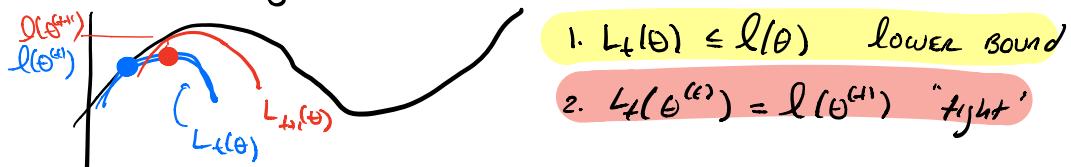


Applications of EM & Factor Analysis

- Finish EM properties
- GAUSSIAN mixture model as EM
- Factor Analysis

RECALL EM Algorithm



Assignment Project Exam Help

LAST TIME

$$l(\theta) = \sum_i \log \sum_z Q_i(z) \frac{P(x^{(i)}, z; \theta)}{Q_i(z)}$$

Add WeChat powcoder $\forall i \sum_z Q_i(z) = 1, Q_i(z) \geq 0$

$$L_t(\theta) = \sum_i \sum_z Q_i(z) \log \frac{P(x^{(i)}, z; \theta)}{Q_i(z)}$$

WE SHOWED Property 1, $l(\theta) \geq L_t(\theta)$ Key STEP IS JENSEN

$$\log \sum_z Q_i(z) \frac{P(x^{(i)}, z; \theta)}{Q_i(z)} \geq \sum_z Q_i(z) \log \frac{P(x^{(i)}, z; \theta)}{Q_i(z)}$$

ELBO($x^{(i)}, z, \theta$)

TO SET Property 2, we pick $Q_i(z) \propto \theta^{(i)}$

$$Q_i(z) = P(z | x^{(i)}; \theta)$$

VERIFY, since $\frac{P(x^{(i)}, z; \theta)}{Q_i(z)} = \frac{P(z | x^{(i)}; \theta)}{P(z | x^{(i)}; \theta)} \text{ does not depend on } z$

$$(LHS) \log \sum_z Q_i(z) \cdot c = \log c$$

(ELBO) $\sum_z Q_i(z) \log c = \log c$. Property 2 holds.

RESTATE EM

$$\begin{aligned}
 (\text{E-STEP}) \quad & \text{for } i=1..n \quad \text{SET } Q_i(z) = P(z^{(i)} | x^{(i)}, \theta^{(t)}) \\
 (\text{M-STEP}) \quad & \theta^{(t+1)} = \underset{\theta}{\operatorname{Argmax}} \mathcal{L}_t(\theta) \\
 & = \underset{\theta}{\operatorname{Argmax}} \sum_{i=1}^n ELBO(x^{(i)}, Q^{(t)}, \theta)
 \end{aligned}$$

DATA & INPUT PARAMS

WARM UP: Mixture of Gaussians. EM RECOVERS OUR AD-HOC ALGORITHM

$$P(x^{(i)}, z^{(i)}) = P(x^{(i)} | z^{(i)}) P(z^{(i)})$$

$$z^{(i)} \sim \text{Multinomial}(1) \quad \phi_i \geq 0 \quad \sum \phi_i = 1 \quad \text{"IN cluster j"}$$

$$x^{(i)} | z^{(i)} = j \sim N(\mu_j, \sigma_j^2) \quad \text{"CLUSTER MEANS"}$$

Assignment Project Exam Help

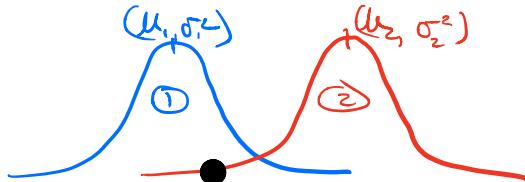
<https://powcoder.com>

WHAT IS EM HERE?

$$Q_i(z) = P(z^{(i)} = j | x^{(i)}; \theta)$$

Add WeChat powcoder

WE SAW THAT COULD COMPUTE VIA BAYES RULE $P(x^{(i)} | z^{(i)} = j; \theta)$



1. much more likely for ① than ②
2. BUT if we knew $\phi_2 \gg \phi_1$, maybe we'd think likely from ②

BAYES RULE AUTOMATES THIS REASONING

M-STEP: Compute Derivatives ..

$$\underset{\phi, \mu, \Sigma}{\operatorname{MAX}} \sum_{i=1}^n \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

WRITE Θ FOR NOTATION ABOVE

$$\text{Rewrite } \omega_j^{(i)} \triangleq Q_i(z=j)$$

$$P(x^{(i)}, z^{(i)}; \Theta) = P(x^{(i)} | z^{(i)}) P(z^{(i)})$$

$$f_i(\Theta) = \sum_j \omega_j^{(i)} \log \left(\frac{1}{2\pi |\Sigma_j|^{1/2}} \exp \left\{ -\frac{1}{2} (x^{(i)} - \mu_j)^T \Sigma_j^{-1} (x^{(i)} - \mu_j) \right\} \cdot \phi_j \right)$$

$$\nabla_{\mu_j} f_i(\Theta) = \sum_i \nabla_{\mu_j} (\omega_j^{(i)} - \frac{1}{2} (x^{(i)} - \mu_j)^T \Sigma_j^{-1} (x^{(i)} - \mu_j))$$

$$= -\frac{1}{2} \sum_i \omega_j^{(i)} \Sigma_j^{-1} (x^{(i)} - \mu_j) = -\frac{1}{2} \sum_j (\sum_i \omega_j^{(i)} (x^{(i)} - \mu_j))$$

Assignment Project Exam Help

SETTING TO 0 AND USING Σ_j IS FULL RANK $\Rightarrow \sum_i \omega_j^{(i)} (x^{(i)} - \mu_j) = 0$

$$\therefore \mu_j = \sum_i \frac{\omega_j^{(i)} x^{(i)}}{\sum_i \omega_j^{(i)}} \quad (\text{AS BEFORE})$$

<https://powcoder.com>

$$\nabla \phi_j = \sum_{i=1}^n \omega_j^{(i)} \nabla_{\phi_j} \log \phi_j + \nabla_{\phi_j} \lambda (\sum_j \phi_j - 1)$$

$$= \sum_{i=1}^n \frac{\omega_j^{(i)}}{\phi_j} + \lambda = 0 \Rightarrow \phi_j = -\frac{1}{\lambda} \sum_{i=1}^n \omega_j^{(i)}$$

$$\text{SINCE } \sum \phi_j = 1, \quad \sum_j \phi_j = -\frac{1}{\lambda} \sum_i \omega_j^{(i)} = -\frac{n}{\lambda}$$

$$\therefore \phi_j = \frac{1}{n} \sum_i \omega_j^{(i)}$$

MESSAGE: EM RECOVERS GMM AUTOMATICALLY.

NB: IF $z^{(i)}$ IS CONTINUOUS, ONE CAN REPLACE SUMS w/ INTERVALS

Factor Analysis

MANY fewer parts than dimensions " $n \ll d$ "

cf: GMMs $n \gg d$ lots of neurons, few sources.

How does this happen?

PLACE SENSORS ALL OVER CAMPUS, RECORD @ 1000s of locations
 $\Rightarrow d \approx 1000s$

But Only record for 30 days ($n < d$)

WANT TO FIT A DENSITY but seems hopeless.

Assignment Project Exam Help

KEY IDEA: Assume there is some latent r.v. that
<https://powcoder.com>
IS NOT TOO complex AND Explains behavior.

Add WeChat powcoder

1st Let's SEE Problems w/ GMMs.... Even 1 GAUSSIAN

$$\mu = \frac{1}{n} \sum_{i=1}^n x^{(i)} \rightarrow \text{this is OK}$$

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

RANK(Σ) $\leq n < d$ - not full rank.

Problem in GAUSSIAN likelihood

$$P(x; \mu, \Sigma) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp \left\{ -(x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

IS NOT DEFINED.

$$\hookrightarrow |\Sigma| = 0$$

WE will fix these issues by examining three models

that are simpler. Spoiler: we'll combine these in the end!

Recall MLE for Gaussian

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

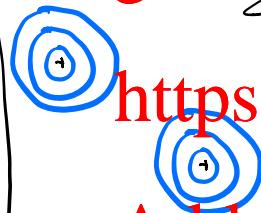
Equivalent

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

If Σ is full rank, $\nabla_\mu = \sum_{i=1}^n \Sigma^{-1} (\mathbf{x}-\mu) = 0 \Rightarrow \mu = \frac{1}{n} \sum_i \mathbf{x}^{(i)}$
we'll use this as plugin below.

Building Block 1

Assignment Project Exam Help
 Suppose INDEPENDENT AND IDENTICAL COVARIANCE
 $\Sigma = \sigma^2 I$ (NB: PARAMETER σ^2)



<https://powcoder.com>

COVARIANCE ARE CIRCLES

Add WeChat powcoder

WHAT IS MLE FOR Σ ?

$$|\Sigma| = 2d$$

$$\underset{\sigma^2}{\text{MIN}} \quad \sigma^{-2} \underbrace{\sum_{i=1}^n (\mathbf{x}^{(i)} - \mu)^T (\mathbf{x}^{(i)} - \mu)}_C + d \log \sigma^2$$

$$\text{let } z = \sigma^2 \quad \underset{z}{\text{MIN}} \quad \frac{1}{z} C + d \log z$$

$$\Rightarrow \frac{1}{z} = -z^{-2} C + \frac{nd}{z} = 0 \Rightarrow z = \frac{C}{nd}$$

$$\therefore \sigma^2 = \frac{1}{nd} \sum_{i=1}^n (\mathbf{x}^{(i)} - \mu)^T (\mathbf{x}^{(i)} - \mu)$$

"SUBTRACT MEAN AND SQUARE ALL ENTRIES."

Building Block 2

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



Axis Aligned ellipse

SET $z_i = \sigma_i^2$ (same idea as above)

$$\min_{z_1 \dots z_d} \sum_{i=1}^n \sum_{j=1}^d z_j^{-1} (x^{(i)} - \mu_j)^2 + \log z_j$$

This is d problems for each 1 dimension

$$\Rightarrow \sum_{i=1}^n z_j^{-1} (x^{(i)} - \mu_j)^2 + \log z_j$$

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

Assignment Project Exam Help

Our FAster model

PARAMETERS

$\mu \in \mathbb{R}^d$

$\Lambda \in \mathbb{R}^{d \times s}$

Add WeChat powcoder

$\Phi \in \mathbb{R}^{d \times d}$ - DIAGONAL MATRIX

Model

$$P(x, z) = P(x|z) P(z) \quad z \text{ IS LATENT}$$

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

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

MEAN IN THE SPACE $\epsilon \sim N(0, \Phi)$ NOISY

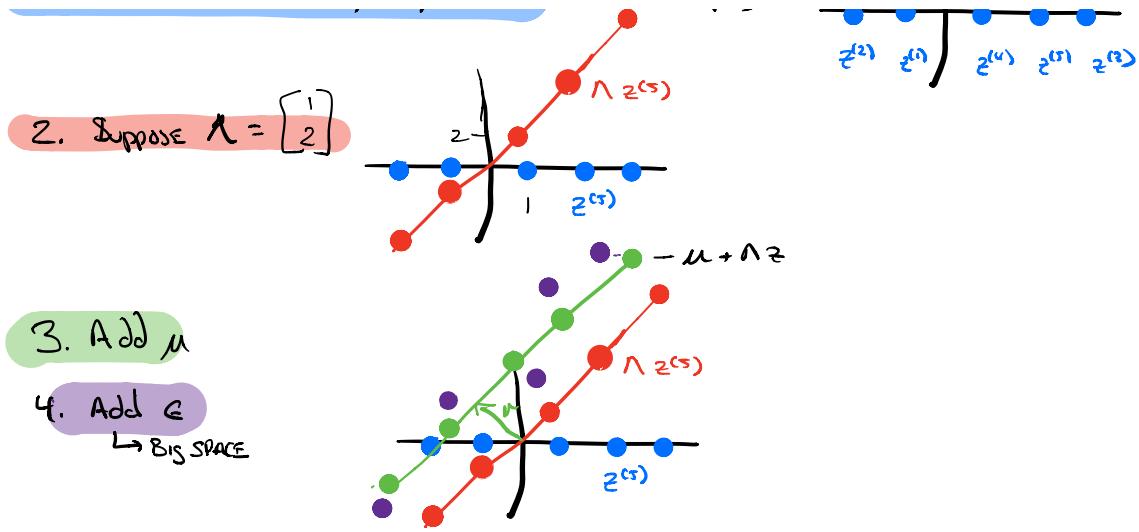
$$\epsilon \sim N(0, \Phi)$$

$$\text{Ex: } d=2, s=1, n=5$$

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

1. GENERATE $z^{(1)}, \dots, z^{(s)}$ from $N(0, I)$





DATA WE WOULD OBSERVE ARE PURPLE DOTS

SO SMALL LATENT SPACE PRODUCES DATA IN HIGH DIM. SPACE.

Assignment Project Exam Help

TECHNICAL TOOLS: Block GAUSSIANS
<https://powcoder.com>

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \quad \begin{array}{l} \mathbf{x}_1 \in \mathbb{R}^{d_1}, \mathbf{x}_2 \in \mathbb{R}^{d_2} \\ \mathbf{x} \in \mathbb{R}^n \end{array}$$

$$\Sigma = \begin{bmatrix} \frac{d_1}{\Sigma_{11}} & \frac{d_2}{\Sigma_{21}} \\ \frac{\Sigma_{12}}{\Sigma_{11}} & \frac{\Sigma_{22}}{\Sigma_{11}} \end{bmatrix}_{1 \times 2}^{1 \times d_1} \quad \Sigma_{ij} \in \mathbb{R}^{d_i \times d_j} \quad i, j \in \{1, 2\}$$

NOTATION IS WIDELY USED AND HELPFUL.

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

FOR GAUSSIANS, $P(x_1) = N(\mu_{11}, \Sigma_{11})$ (NOT SWAPPING)

FACT 2: $P(x_1 | x_2) \sim N(\mu_{1|2}, \Sigma_{1|2})$ CONDITIONING

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

$$\hat{\Sigma}_{12} = \hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \quad (\text{matrix inversion lemma})$$

Proofs outline (apply to AdB)

Summary: Marginalization $\not\equiv$ Conditioning Gaussian \Rightarrow
 Another GAUSSIAN (CLOSED)
 WE HAVE formula for PARAMETERS.

Back to Factor Analysis

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

$\begin{pmatrix} z \\ x \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \Sigma \right)$ SINCE $E[z] = 0$
 $E[x] = \mu$

WHAT IS Σ ?
 $\Sigma_{11} = E[zz^T] = \Lambda$

$$\begin{aligned} \hat{\Sigma}_{12} &= E[z(x-\mu)^T] = E[z z^T \Lambda^T] + E[z \epsilon^T] \\ &= \Lambda^T \end{aligned}$$

$$\hat{\Sigma}_{21} = \hat{\Sigma}_{12}^T$$

$$\begin{aligned} \hat{\Sigma}_{22} &= E[(x-\mu)(x-\mu)^T] \\ &= E[(\Lambda z + \epsilon)(\Lambda z + \epsilon)^T] \\ &= E[\Lambda z z^T \Lambda^T] + E[\epsilon \epsilon^T] \\ &= \Lambda \Lambda^T + \Phi \end{aligned}$$

$$\hat{\Sigma} = \begin{bmatrix} \Lambda & \Lambda^T \\ \Lambda & \Lambda \Lambda^T + \Phi \end{bmatrix}$$

E-STEP : $Q_i(z) = P(z^{(i)} | x^{(i)}; \theta)$ - USE CONDITIONAL!

M-STEP : WE HAVE CLOSED FORMS!

Summary:

- WE SAW THAT EM CAPTURES GMM
- WE LEARNED ABOUT FACTOR ANALYSIS (Latent low dim. STRUCTURE)
- WE SAW HOW TO ESTIMATE PARAMETERS OF FA USING EM.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder