CO2 774
Machine Learning
Oct 23, 2021

(1) Post # 49

Last class:-

K-Means:- first unsupervised algorithm (clustering)

↳ GMM:- Density estimation (Gaussian Mixture Models)

$\{x^{(i)}\}_{i=1}^{m}$ $\quad x^{(i)} \in \mathbb{R}^n$

↳ $\{x^{(i)}, z^{(i)}\}_{i=1}^{m}$ ↳ hidden $\quad | \quad \{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

$z^{(i)} \in \{1 \ldots r\}$

$\prod_{i=1}^{m} P(x^{(i)}; \theta)$ = Likelihood function

$P(x^{(i)}, z^{(i)}; \theta)$ ←

↳ $P(x^{(i)} | z^{(i)}; \theta) \sim N(\mu_{z^{(i)}}, \Sigma_{z^{(i)}})$

↳ $P(z^{(i)}; \theta) \sim \text{Multinomial}(\phi)$

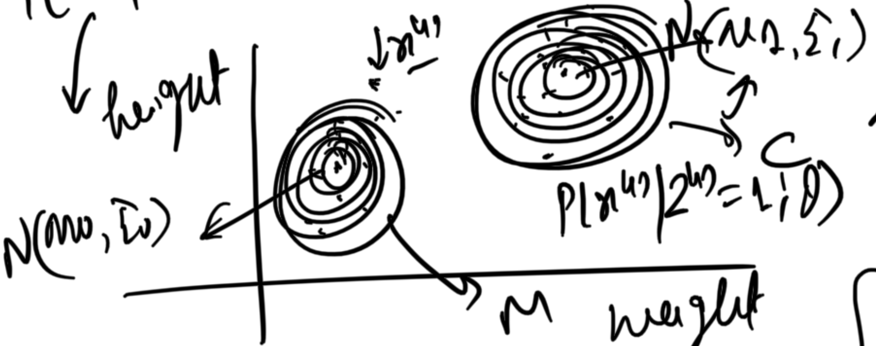
$\underset{\theta}{\arg\max} \log \prod_{i=1}^{m} P(x^{(i)}; \theta)$ $\quad \{x^{(i)}\}_{i=1}^{m}$

$\Downarrow$

$\sum_{z^{(i)}} P(x^{(i)}, z^{(i)}; \theta)$ →

$\underset{\theta}{\arg\max} \log \prod_{i=1}^{m} P(x^{(i)}, y^{(i)}; \theta)$ $\quad \{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

$P(x^{(i)} | z^{(i)} = 0; \theta)$



Labels are hidden

$\{x^{(i)}\}_{i=1}^{m}$

$y^{(i)} \in \{M, C\}$

$\{0, 1\}$

$z^{(i)} \in \{0, 1\}$

$z^{(i)} \in \{1 \ldots r\}$ $\quad$ $[y^{(i)} \in \{1 \ldots r\}]$

$$\left.\begin{aligned} p(x^{(i)} \mid z^{(i)}=0;\theta) &= \mathcal{N}(\mu_0, \Sigma_0) \\ p(x^{(i)} \mid z^{(i)}=1;\theta) &= \mathcal{N}(\mu_1, \Sigma_1) \end{aligned}\right] \text{ In general, can have } r \text{ classes}$$

But we don't know $z^{(i)}$s:- $\quad x^{(i)} \in \mathbb{R}^n$

$$\log \prod_{i=1}^{m} p(x^{(i)};\theta) = \log \prod_{i=1}^{m} \sum_{z^{(i)}} p(x^{(i)} \mid z^{(i)};\theta)\, p(z^{(i)};\theta)$$

$z^{(i)} \in \{1 \ldots r\}$

✓ $z^{(i)} \sim \text{Multinomial}(\Phi)$ $\quad$ $\Phi = (\phi_1 \ldots \phi_r)$

✓ $x^{(i)} \mid z^{(i)} \sim \mathcal{N}(\mu_{z^{(i)}}, \Sigma_{z^{(i)}})$ $\quad$ $\sum \phi_k = 1$

Parameters:-

$$\Theta = \left(\Phi, \{\mu_k, \Sigma_k\}_{k=1}^{r}\right)$$

GMM

$$\arg\max_{\theta} \log\left[\prod_{i=1}^{m} p(x^{(i)};\theta)\right]$$

$$\sum_{z^{(i)}} p(x^{(i)} \mid z^{(i)};\theta)\, p(z^{(i)};\theta)$$

GDA

$$\arg\max_{\theta} \log\left[\prod_{i=1}^{m} p(x^{(i)}, z^{(i)};\theta)\right]$$

Gaussian: $p(x^{(i)} \mid z^{(i)};\theta) = \mathcal{N}(\mu_{z^{(i)}}, \Sigma_{z^{(i)}})$

Mixture:- ~~Mix~~ from a set of all components

MIXTURE MODELS $\quad$ objective:- Find θ s.t $LL(\theta)$ is maximized

$$\hookrightarrow \arg\max_{\theta} LL(\theta) = \arg\max_{\theta} \log\left[\prod_{i=1}^{m} p(x^{(i)};\theta)\right] \quad \boxed{p(x^{(i)};\theta)}$$

θ

$$= \arg\max_{\theta} \sum_{i=1}^{m} \log P(x^{(i)}; \theta)$$

$$LL(\theta) = \arg\max_{\theta} \sum_{i=1}^{m} \log \sum_{z^{(i)}} \left[P(x^{(i)} \mid z^{(i)}; \theta)\, P(z^{(i)}; \theta) \right]$$

In GDA:- $\sum_{z^{(i)}}$ was not there

How to solve this optimization problem?

↳ Idea:- Differentiate: Gradient descent

So: An algorithm similar to K-Means:-

GMM

(A) If we know $\theta = (\Phi, \{\mu_k, \Sigma_k\}_{k=1}^{r})$ then we can estimate $z^{(i)}$'s

E-step $P(z^{(i)} \mid x^{(i)}; \theta) = \dfrac{P(x^{(i)} \mid z^{(i)}; \theta)\, P(z^{(i)}; \theta)}{P(x^{(i)}; \theta)}$

$= Q_i(z^{(i)})$

(soft)

K-Means

(A) If $(\mu_1 \ldots \mu_r)$ are known then we can estimate $c^{(i)}$'s

E-step (Hard)

(B) If we knew $c^{(i)}$'s then we can estimate $(\mu_1 \ldots \mu_r)$

(B) If we knew $z^{(i)}$'s $(Q_i(z^{(i)}))$, we can estimate θ

$1\{z^{(i)} = k\}$ → $\{\sum Q_i(z^{(i)})\}$

⇓ same as GDA

$$\arg\max_{\theta} \left[\sum_{i=1}^{m} \log P(x^{(i)} \mid z^{(i)}; \theta)\, P(z^{(i)}; \theta) \right]$$

Analytically

Problem:- $z^{(i)}$'s are specified using a distribution (GMM)

Estimate of parameters:- → GDA $1\{y^{(i)} = k\}$

$$\phi_k = \frac{\sum_{i=1}^{m} Q_i(z^{(i)}=k)}{m} \rightarrow 1\{y^{(i)}=k\}$$

$$\mu_k = \frac{\sum_{i=1}^{m} Q_i(z^{(i)}=k)\, x^{(i)}}{\sum_{i=1}^{m} Q_i(z^{(i)}=k)} \rightarrow 1\{y^{(i)}=k\}$$

$$\Sigma_k = \frac{\sum_{i=1}^{m} Q_i(z^{(i)}=k)(x^{(i)}-\mu_k)(x^{(i)}-\mu_k)^T}{\sum_{i=1}^{m} Q_i(z^{(i)}=k)}$$

Final Algorithm:- $\{x^{(i)}\}_{i=1}^{m}$

↳ $z^{(i)} \sim \text{Multinoulli}(\Phi)$

$x^{(i)}|z^{(i)} \sim N(\mu_{z^{(i)}}, \Sigma_{z^{(i)}})$

$\Theta \leftarrow \text{init}();$

do{

(A) Estimate $Q_i(z^{(i)})$ $[P(z^{(i)}|x^{(i)};\theta)]$ $\forall i$ [E-step]

(B) find $\arg\max_{\theta} \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \underbrace{[P(x^{(i)},z^{(i)};\theta)]}_{\text{M-step}}$ $\rightarrow$ $P(x^{(i)}|z^{(i)};\theta)\,P(z^{(i)};\theta)$

} (while ! converged)

convergence :-

EM:- Expectation Maximization :-> known

$\{x^{(i)}\}_{i=1}^{m} \longrightarrow \underline{P(x^{(i)}, z^{(i)}; \theta)}$

hidden

$\log \prod_{i=1}^{m} P(x^{(i)}; \theta)$: log likelihood

We formally prove that (GMM like) (E-step – M-step) iterative algorithm converges to local optima of log likelihood

(Block coordinate) $d_i(z^{(i)})\ (\theta)$