Lecture 14 Expectation-Maximization Algorithms

< Now starts unsupervised learning >

1. K-Means Algorithm (Clustering)

Data: $\{x^{(1)}, \cdots, x^{(m)}\}$

① Initialize cluster centroids: $\mu_1 \cdots \mu_k \in \mathbb{R}^n$ randomly

② Repeat until convergence:

L_2 norm ($||x||$ or $||x||_2$); l_1 norm ($||x||_1$)

a. Set $c^{(i)} = \arg\min_j ||x^{(i)} - \mu_j||_{\boxed{2}}$ (assign points to the nearest centroid)

b. For $j = 1, \cdots, k$,

$$\mu_{ji} = \sum_{i=1}^{m} I\{c^{(i)} = j\} x^{(i)} / \sum_{i=1}^{m} I\{c^{(i)} = j\}$$ (move the clusters centroids)

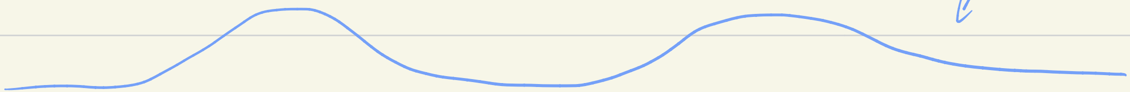
assignment centroids

- Cost Function: $J(c, \mu) = \sum_{i=1}^{m} ||x^{(i)} - \mu_{c^{(i)}}||^2$
- How to select k? — prefer manually

2. Mixture of Gaussian Model

Overall distribution:

two gaussian



But! Don't have label! Don't know which example comes from which Gaussian

Suppose there's a latent (hidden / unobserved) random variable z,
and $x^{(i)}, z^{(i)}$ are distributed $P(x^{(i)}, z^{(i)}) = P(x^{(i)} | z^{(i)}) \cdot P(z^{(i)})$,
where $z^{(i)} \sim \text{Multinomial}(\phi)$, $x^{(i)} | z^{(i)} = j \sim N(\mu_j, \Sigma_j)$

If only two classes, then $z^{(i)} \sim$ Bernoulli

- If we know $z^{(i)}$'s, we can use MLE:

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log p(x^{(i)}, z^{(i)}; \phi, \mu, \Sigma)$$

$$\phi_j = \frac{1}{m} \sum_{i=1}^{m} I\{z^{(i)} = j\}, \quad \mu_j = \frac{\sum_{i=1}^{m} I\{z^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} I\{z^{(i)} = j\}}$$

⇒ Use EM (Expectation Maximization)

① E-step (Guess value of $z^{(i)}$s)

$$\text{Set } w_j^{(i)} = P(z^{(i)}=j \mid x^{(i)}; \phi, \mu, \Sigma)$$

$$= \frac{\boxed{P(x^{(i)} \mid z^{(i)}=j)} \cdot \boxed{P(z^{(i)}=j)}}{\sum_{l=1}^{k} P(x^{(i)} \mid z^{(i)}=l)\, P(z^{(i)}=l)} \rightarrow \sim \text{Multinomial}(\phi) \Rightarrow \phi_j$$

$\sim N(\mu_j, \Sigma_j) \Rightarrow \dfrac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_j)^T \Sigma^{-1} (x^{(i)}-\mu_j)\right)$

② M-step

$$\phi_j = \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)} \; ; \quad \mu_j = \frac{\sum_{i=1}^{m} w_j^{(i)} \cdot x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}$$

Replace $\sum_{i=1}^{m} I\{z^{(i)}=j\}$ with $w_j^{(i)}$

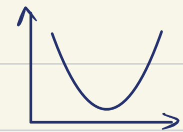
⇒ $\because w_j^{(i)} = E[I\{z^{(i)}=j\}]$

⇒ Intuition: ① a softer way to assign points to different classes, uses possibility as weighting.

② can fit a mixture of more than two Gaussians and richer model

* Can also do anomaly test: $\begin{cases} p(x) \geq \varepsilon & \text{ok} \\ p(x) < \varepsilon & \text{anomaly} \end{cases}$

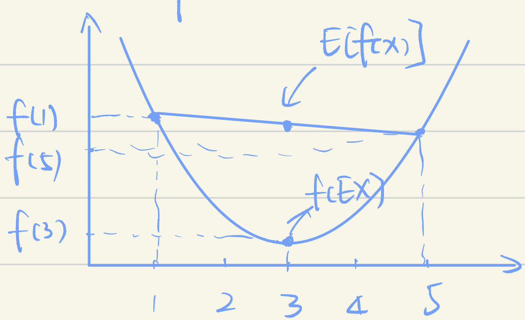
3. Jensen's inequality



Let f be a convex function (e.g. $f(x)'' > 0$)

Let X be a random variable, then $f(EX) \leq E(f(x))$

<example>:



if $X = \begin{cases} 1 & \text{with prob } \frac{1}{2} \\ 5 & \text{with prob } \frac{1}{2} \end{cases}$

Further, if $f''(x) > 0$ (f is strictly convex),

then $E[f(x)] = f(EX) \iff X$ is a constant.

* If concave, then everything flipped around the other way.

4. EM Algorithm

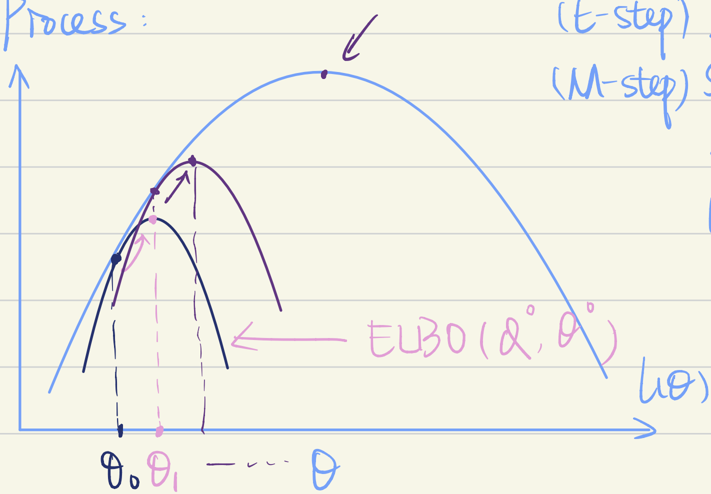
Have model for $P(x, z; \theta)$

Only observe x. $\{x^{(1)}, x^{(2)}, \cdots x^{(m)}\}$

$$l(\theta) = \sum_{i=1}^{m} \log P(x^{(i)}, \theta) = \sum_{i=1}^{m} \log \sum_{z^{(i)}} P(x^{(i)}, z^{(i)}; \theta)$$

Want: $\arg\max_\theta l(\theta)$

Process:



(E-step) Step 1: Create a lower bound of θ_0.

(M-step) Step 2: find the max θ of the lower bound

Repeat step 1 and step 2 until find the local maximum θ

E-step, lower bound: $l(\theta)$

$$\Rightarrow \max_\theta \sum_i \boxed{\log P(x^{(i)};\theta)} = \sum_i \log \sum_{z^{(i)}} P(x^{(i)}, z^{(i)};\theta)$$

(evidence lower bound)
ELBO$(x; Q, \theta)$

$$= \sum_i \log \sum_{z^{i}} Q_i(z^{(i)}) \left[\frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}\right]$$

where $Q_i(z^{(i)})$ is a probability distribution $\Rightarrow \sum_{z^{(i)}} Q_i(z^{(i)}) = 1$

$$= \sum_i \log E_{z^{(i)}\sim Q_i}\left[\frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}\right] \geq \sum_i \underbrace{\boxed{E_{z^{(i)}\sim Q_i}\left[\log \frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}\right]}}_{\text{E-step, lower bound: } l(\theta)}$$

(Jensen's inequality concave form)

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

On a given iteration at EM (with parameter θ), we want:

$$\log E_{z^{(i)}\sim Q_i}\left[\frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}\right] = E_{z^{(i)}\sim Q_i}\left[\log \frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}\right]$$

$\Rightarrow \dfrac{P(x^{(i)}, z^{(i)})}{Q_i(z^{(i)})}$ is constant $\Rightarrow$ Set $Q_i(z^{(i)}) \propto P(x^{(i)}, z^{(i)})$ — a proportion of

$$\because \sum_{z^{(i)}} Q_i(z^{(i)}) = 1 \Rightarrow Q_i(z^{(i)}) = \frac{P(x^{(i)}, z^{(i)};\theta)}{\sum_{z^{(i)}} P(x^{(i)}, z^{(i)};\theta)} \cdots = P(z^{(i)} \mid x^{(i)};\theta)$$

* Summary:

E-step: Set $Q_i(z^{(i)}) = P(z^{(i)} \mid x^{(i)};\theta)$

$w_j^{(i)}$

M-step: $\theta := \arg\max_{\theta} \sum_{i} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$