# CS 307 Lecture 26

Training set $S = \{x^{(1)}, x^{(2)}, \dots x^{(n)}\}$.

**Clustering** : Group data in the training set into clusters.

Training set $S = \{x^{(1)} \dots x^{(n)}\}$.

Goal : Discover the distribution of the training data.


Distribution of the training data is most conveniently described using "latent"/hidden variables.

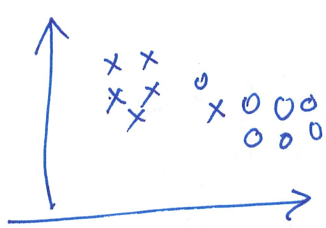
"Mixture of Gaussians" : $x \in \mathbb{R}^d$ ←

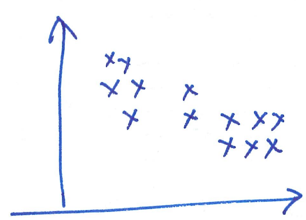
- Pick $z \in \{1, \dots k\}$ is distributed according to multinomial distribution $\phi$.

$$\phi_i = P[z = i]$$

- $x \sim \mathcal{N}(\mu_z, \Sigma_z)$

$$p(x|z) = \frac{1}{(2\pi)^{d/2}|\Sigma_z|^{1/2}} \cdot e^{\frac{1}{2}\left((x-\mu_z)^T \Sigma_z^{-1} (x-\mu_z)\right)}$$





GDA

Given a training set $S$

Goal: Discover the parameters that define mixture of Gaussians distribution from which $S$ is drawn.

$$\left(\phi, \{\mu_z, \Sigma_z\}_{z=1}^{k}\right)$$

$S = \{ x^{(1)} \ldots x^{(n)} \}$.

$$\text{pdf}\,(S;\ \phi, \mu_1, \Sigma_1, \ldots \mu_k, \Sigma_k) = \prod_{i=1}^{n} p(x^{(i)};\ \phi, \mu, \Sigma)$$

$$L(S; \phi, \mu_j, \Sigma_j) = \sum_{i=1}^{n} \log\, p(x^{(i)};\ \phi, \mu_j, \Sigma_j)$$

$$= \sum_{i=1}^{n} \log \sum_{z=1}^{k} p(x^{(i)}, z;\ \phi, \mu_j, \Sigma_j)$$

$$= \sum_{i=1}^{n} \log \sum_{z=1}^{k} \underbrace{p(z;\ \phi, \mu_j, \Sigma_j)}_{\text{Multinomial}(\phi)}\ \underbrace{p(x^{(i)} \mid z;\ \phi, \mu_j, \Sigma_j)}_{\mathcal{N}(\mu_z, \Sigma_z)\ \text{Gaussian}}$$

---

Suppose $z^{(i)}$ is known ($z^{(i)}$ is the distribution from which $x^{(i)}$ is drawn)

then find the parameters that maximize likelihood is easy.

$$p(z=j) = \phi_j = \frac{\sum_{i=1}^{n} 1[z^{(i)} = j]}{n}$$

$$\mu_j = \frac{\sum_{i=1}^{n} 1[z^{(i)} = j]\ x^{(i)}}{\sum_{i=1}^{n} 1[z^{(i)} = j]}$$

$$\Sigma_j = \frac{\sum_{i=1}^{n} 1[z^{(i)} = j]\ (x-\mu_j)(x-\mu_j)^T}{\sum_{i=1}^{n} 1[z^{(i)} = j]}$$

---

$$w_j^{(i)} = p(z=j \mid x^{(i)};\ \phi \ldots)$$

$$= \frac{p(z=j)\ p(x^{(i)} \mid z=j;\ \phi \ldots)}{p(x^{(i)};\ \phi \ldots)}$$

$$= \frac{\phi_j\ p_{N(\mu_j, \Sigma_j)}(x^{(i)})}{\sum_{l=1}^{k} p(x^{(i)}, z=l;\ \phi \ldots)} = \frac{\phi_j\ p_{N(\mu_j, \Sigma_j)}(x^{(i)})}{\sum_{l=1}^{k} \phi_l\ p_{N(\mu_l, \Sigma_l)}(x^{(i)})}$$

## → Expectation - Maximization for Mixture of Gaussians:

Initialize $\phi, \mu_1, \Sigma_1 \ldots \mu_k \Sigma_k$

Repeat

E-step:

$$w_j^{(i)} = p(z=j \mid x^{(i)}; \phi, \mu_1, \Sigma_1, \ldots \mu_k, \Sigma_k)$$

M-step:

$$\phi_j = \frac{1}{n}\sum_{i=1}^{n} w_j^{(i)}$$

$$\mu_j = \frac{\sum_{i=1}^{n} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{n} w_j^{(i)}}$$

$$\Sigma_j = \ldots$$

Until convergence.

▷ EM algorithm find ~~a~~ local optimum for the parameters.

EM algorithm is a more general approach to maximizing likelihood when the likelihood has the form

$$\sum \log \sum \ldots$$