$\boxed{\text{Gaussian Mixture Models}}$

$\boxed{\text{Linear Regression}}$

(Discriminative) $P(y^{(i)} | x^{(i)}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(y^{(i)} - w^T\phi(x^{(i)})\right)^2\right)$

$\underbrace{\text{(generative)}}_{\text{no labels}}$: $P(x^{(i)} | \mu, \sigma) = \mathcal{N}(x^{(i)} | \mu, \sigma)$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^{(i)} - \mu)^2\right)$$

We want to fit out pts to $\mathcal{N}$

Given $D = \{(x^{(i)})\}$ find $\theta = \{\mu, \sigma\}$ of

$$P(x^{(i)} | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^{(i)} - \mu)^2\right)$$

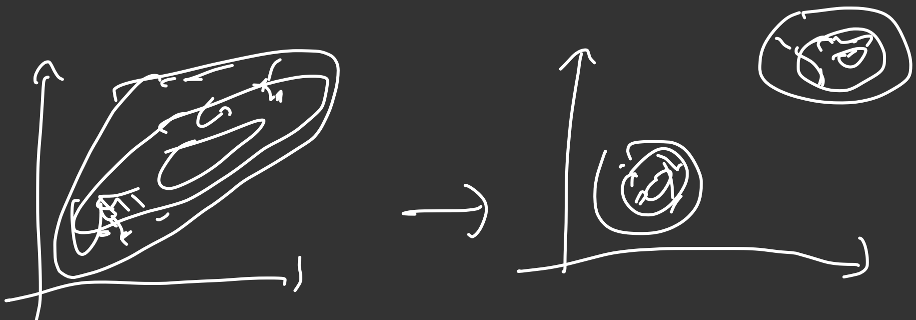
$\boxed{\text{MLE}} \rightarrow \min_{\mu, \sigma} -\log \prod P(x^{(i)} | \mu, \sigma)$

$$= \sum_{i \in D} \frac{1}{2\sigma^2}(x^{(i)} - \mu)^2 + \frac{N}{2}\log(2\pi\sigma^2)$$

$$\frac{\partial}{\partial \mu} : = 0 \implies \mu = \frac{1}{N} \sum_{i \in D} x^{(i)}$$

$$\frac{\partial}{\partial \sigma} : = 0 \implies \sigma^2 = \frac{1}{N} \sum_{i \in D} (x^{(i)} - \mu)^2$$

But: $\boxed{\text{fit multiple gaussians.}}$



$$\Downarrow$$

$$P(x^{(i)} \mid \pi, \mu, \sigma) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x^{(i)} \mid \mu_k, \sigma_k)$$

$$\begin{cases} \sum_k \pi_k = 1 \quad \text{s.t. } \pi_k \geq 0 \ \forall k \\ \min_{\pi, \mu, \sigma} -\log \prod_{i \in D} P(x^{(i)} \mid \pi, \mu, \sigma) \end{cases}$$

$\boxed{\text{latent}}$

$z_{ik} \in \{0, 1\}$ with $\sum_{k=1}^{K} z_{ik} = 1 \ \forall i$

$p(z_{ik} = 1) = \pi_k$ $\qquad p(z_i) = \prod_{k=1}^{K} \pi_k^{z_{ik}}$

$z_i = [z_{i1} \dots z_{iK}]^T$

$= \pi_k$ for which $z_{ik} = 1$

$\underbrace{p(x^{(i)} \mid z_{ik} = 1)}_{}$

ith pt in cluster k

$= N(x^{(i)} \mid \mu_k, \sigma_k)$

$p(x^{(i)} \mid \pi, \mu, \sigma) =$

$\sum_{z_i} p(x^{(i)} \mid z_i)\, p(z_i) = \sum_{z_i} \prod_{k=1}^{K} \pi_k^{z_{ik}} N(x^{(i)} \mid \mu_k, \sigma_k)^{z_{ik}}$

$= \sum_{k=1}^{K} \pi_k N(x^{(i)} \mid \mu_k, \sigma_k)$

$$r_{ik} = P(z_{ik}=1 \mid x^{(i)})$$

$$= \frac{P(z_{ik}=1)\,P(x^{(i)} \mid z_{ik}=1)}{\sum_{\hat{k}=1}^{K} P(z_{i\hat{k}}=1)\,P(x^{(i)} \mid z_{i\hat{k}}=1)}$$

$$= \frac{\pi_k N(x^{(i)} \mid \mu_k, \sigma_k)}{\sum_{\hat{k}=1}^{K} \pi_{\hat{k}} N(x^{(i)} \mid \mu_{\hat{k}}, \sigma_{\hat{k}})}$$

$$\min_{\pi,\mu,\sigma} -\log P(X^{(i)} \mid \pi, \mu, \sigma)$$

$$= -\sum_{i \in D} \log \sum_{k=1}^{K} \pi_k N(x^{(i)} \mid \mu_k, \sigma_k)$$

$$\text{s.t.} \sum_{k=1}^{K} \pi_k = 1$$

$$\frac{\partial}{\partial \mu_k}: \quad \mu_k = \frac{1}{N_k} \sum_{i \in D} r_{ik} x^{(i)}$$

$$\frac{\partial}{\partial \sigma_k}: \quad \sigma_k^2 = \frac{1}{N_k} \sum_{i \in D} r_{ik} (x^{(i)} - \mu_k)^2$$

$$\frac{\partial}{\partial \pi_k}: \quad \sum_{i \in D} \frac{N(x^{(i)} \mid \mu_k, \sigma_k)}{\sum_{k=1}^{K} \pi_k N(x^{(i)} \mid \mu_k, \sigma_k)} + \lambda = 0$$

$$\pi_k = -\frac{N_k}{N}$$

given r_{ik}

M

$$r_{ik} = \frac{\pi_i \mathcal{N}(x^{(i)} | \mu_k, \sigma_k)}{\sum_{\hat{k}} \pi_{\hat{k}} \mathcal{N}(x^{(i)} | \mu_{\hat{k}}, \sigma_{\hat{k}})}$$ given μ_k, π_k, σ_k

$\boxed{E}$

~~closed~~ form → complicate

GD → LR,

Q: ① why latent ? org no closed form.

② EM ? r_{ik} $\pi \mu \sigma$, alternatively update.

③ Lagrange (how) $\min f(x)$ given $g(x) = k$ $\nabla f = \lambda \nabla g$, $g = k$

Gaussian Mixture Model

Initialize μ, σ, π

Iterate: E-step: update

$$r_{ik} = \frac{\pi_k \mathcal{N}(x^{(i)} |}{\sum_{\hat{k}=1}^{k} \pi_{\hat{k}} \mathcal{N}(x^{(i)} | \mu_{\hat{k}}, \sigma_{\hat{k}}^2)}$$

M-step:

$$\mu_k = \frac{1}{N_k} \sum_{i \in D} r_{ik} x^{(i)}$$

$$\sigma_k^2 = \frac{1}{N_k} \sum_{i \in D} r_{ik} (x^{(i)} - \mu_k)^2$$

$$\pi_k = \frac{N_k}{N}$$

$\boxed{\text{Change mixture Gaussian to K-means}}$

$\sigma_k^2 = \epsilon \quad \forall k$

$$\Rightarrow \quad r_{ik} = \frac{\pi_k \exp\left(-\frac{1}{2\epsilon}(x^{(i)} - \mu_k)^2\right)}{\sum_{\hat{k}=1}^{K} \pi_{\hat{k}} \exp\left(-\frac{1}{2\epsilon}(x^{(i)} - \mu_{\hat{k}})^2\right)}$$

$\Rightarrow \quad \epsilon \to 0$

$\Rightarrow$ All responsibility goes to 0

except the one for which

$(x^{(i)} - \mu_{\hat{k}})^2$ smallest

$\Rightarrow \quad r_{ik} = \{0, 1\}$