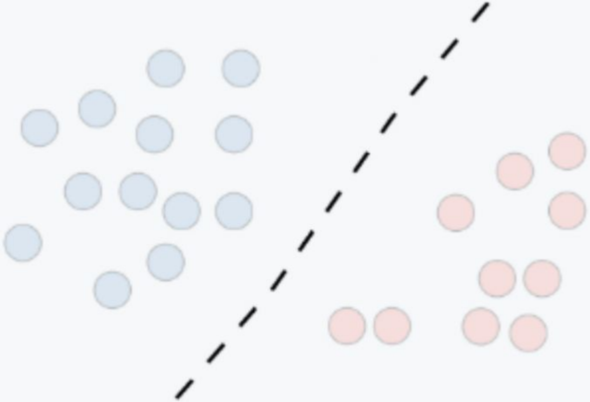
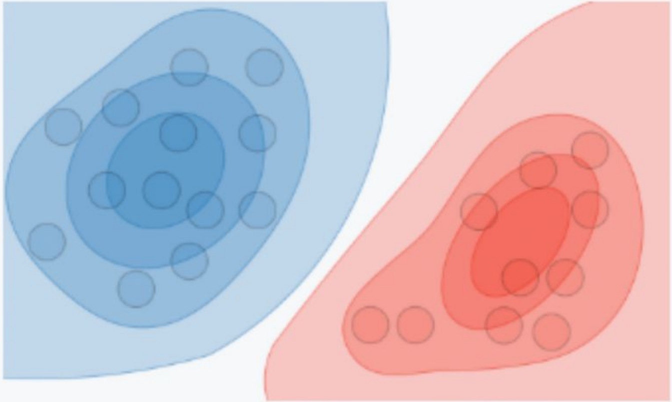
Advanced Machine Learning

Likhit Nayak

Discriminative vs Generative models

	Discriminative model	Generative model
Goal	Directly estimate $P(y\|x)$	Estimate $P(x\|y)$ to then deduce $P(y\|x)$
What's learned	Decision boundary	Probability distributions of the data
Illustration		
Examples	Regressions, SVMs	GDA, Naive Bayes

Maximum Likelihood Estimation (MLE)

Likelihood function:

Joint probability of observed data viewed as a function of the parameters of a statistical model

$$L(\theta|x_1, x_2, ...x_n) = f(x_1, x_2, ...x_n|\theta)$$

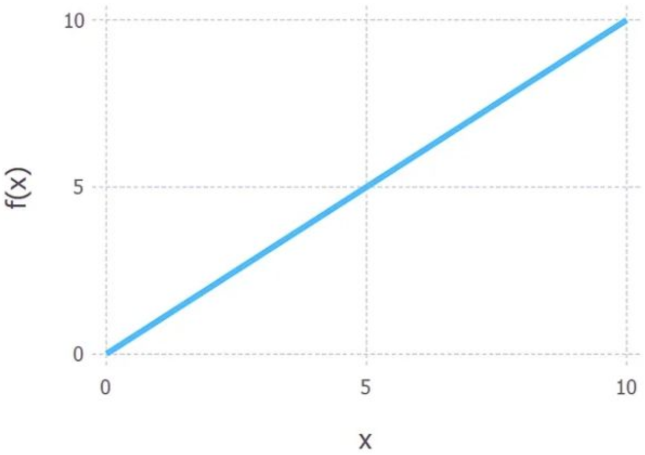
Maximum Likelihood Estimation (MLE)

If we assume all the observed data to be **independently and identically distributed, or i.i.d,** then the cumulative likelihood considering all data points is a product of individual likelihoods.

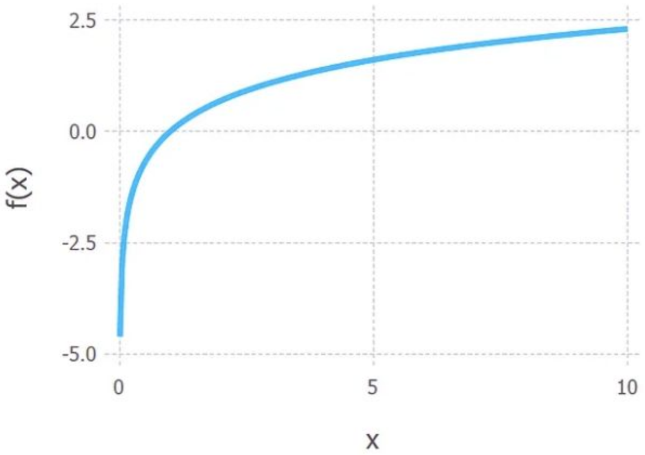
$$\mathcal{L}_n(X_1, X_2, \cdots, X_n, \theta) = \prod_{i=1}^{n} p_\theta(x_i)$$

Maximum Likelihood Estimation (MLE)

Log-likelihood: Taking the log of the likelihood function



(a) $f(x) = x$



(b) $f(x) = \ln(x)$

Maximum Likelihood Estimation (MLE)

MLE finds the parameters that maximizes the log-likelihood function

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max} \ln \mathcal{L}_n(\theta\,; \mathbf{y})$$

Latent variable in GMMs

z*$_i$* is a latent variable that denotes the Gaussian cluster ID for datapoint ***x****$_i$*

$$\mathbb{P}\left(z^{(i)} = j\right) = \pi_j$$

$$\mathbb{P}\left(x^{(i)} \middle| z^{(i)} = j; \theta\right) = \mathcal{N}(x^{(i)} | \mu_j, \Sigma_j)$$

$$\mathbb{P}\left(x^{(i)} \middle| \theta\right) = \sum_{j=1}^{K} \mathbb{P}\left(z^{(i)} = j\right) \cdot \mathbb{P}\left(x^{(i)} \middle| z^{(i)} = j; \theta\right)$$

Expectation Maximization of GMM

$$\sum_{i=1}^{n} \log(p(x_i)) = \sum_{i=1}^{n} \log(\sum_{j=1}^{k} \pi_j \phi(x_i; \mu_j, \Sigma_j)).$$

If we knew the latent variable $\boldsymbol{z_i}$ for all the data points $\boldsymbol{x_i}$

$$= \sum_{i=1}^{n} (\log(p(x_i|z_i)) + \log(p(z_i)))$$

$$= \sum_{i=1}^{n} (\log(\phi(x_i; \mu_{z_i}, \Sigma_{z_i})) + \log(\pi_{z_i}))$$

Expectation Maximization of GMM

$$\sum_{i=1}^{n} \log(p(x_i)) = \sum_{i=1}^{n} \log(\sum_{j=1}^{k} \pi_j \phi(x_i; \mu_j, \Sigma_j)).$$

(E-step) [Expectation step]: Compute soft class memberships, given the current parameters:

$$\tau_{ij} = P(z_i = j | x_{ij}, \pi, (\mu_\ell, \Sigma_\ell)).$$

Expectation Maximization of GMM

(M-step) [Maximization step]: Update parameters by plugging in τ_{ij} (our guess) for the unknown $\mathbb{I}[z_i = j]$, which gives us:

$$\pi_j = \frac{1}{n} \sum_{i=1}^{n} \tau_{ij}, \quad \mu_j = \frac{\sum_{i=1}^{n} \tau_{ij} x_i}{\sum_{i=1}^{n} \tau_{ij}},$$

$$\Sigma_j = \frac{\sum_{i=1}^{n} \tau_{ij}(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{n} \tau_{ij}}.$$