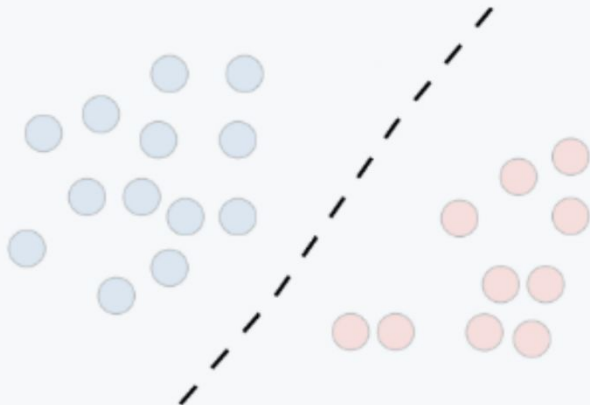
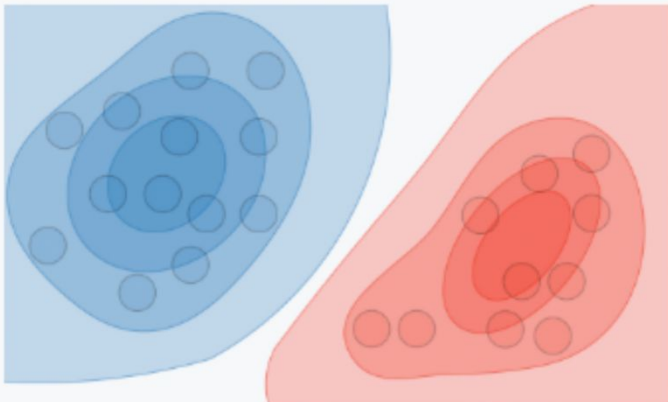


Advanced Machine Learning

Likhith 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	 A scatter plot illustrating a discriminative model. It shows two classes of data points: blue circles on the left and red circles on the right. A dashed diagonal line represents the decision boundary separating the two classes.	 A scatter plot illustrating a generative model. It shows two classes of data points: blue circles on the left and red circles on the right. Each class is enclosed by a shaded region representing its probability distribution. The blue region is on the left and the red region is on the right, with some overlap between them.
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

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

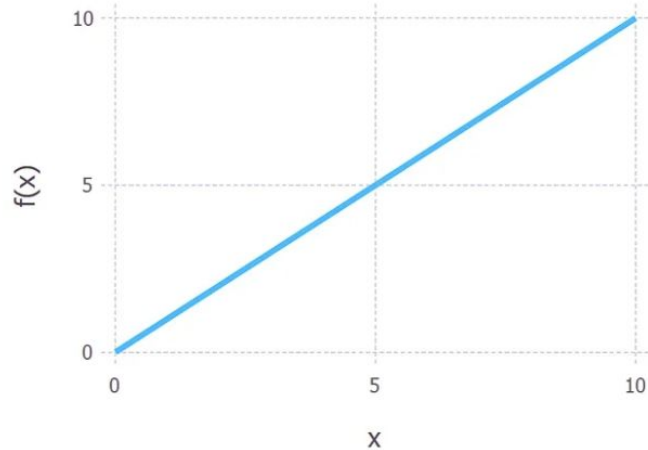
Maximum Likelihood Estimation (MLE)

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.

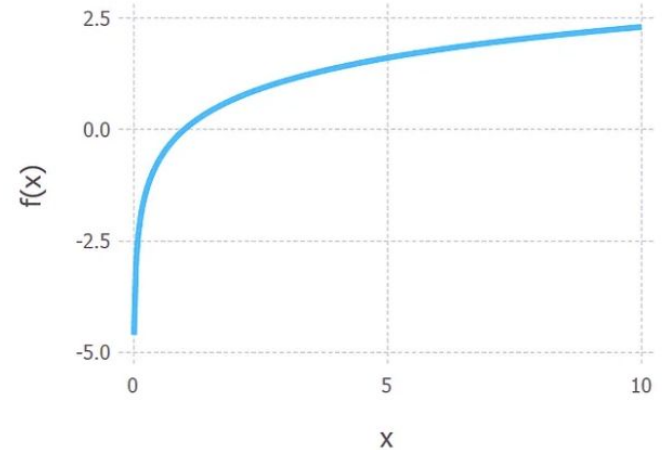
$$\mathcal{L}_n(X_1, X_2, \dots, X_n, \theta) = \prod_{i=1}^n p_{\theta}(x_i)$$

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

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

Latent variable in GMMs

z_i is a latent variable that denotes the Gaussian cluster ID for datapoint \mathbf{x}_i

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

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

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

Expectation Maximization of GMM

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

If we knew the latent variable \mathbf{z}_i for all the data points \mathbf{x}_i

$$\begin{aligned} &= \sum_{i=1}^n (\log(p(x_i|z_i)) + \log(p(z_i))) \\ &= \sum_{i=1}^n (\log(\phi(x_i; \mu_{z_i}, \Sigma_{z_i})) + \log(\pi_{z_i})) \end{aligned}$$

Expectation Maximization of GMM

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

(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:

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

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