Classification

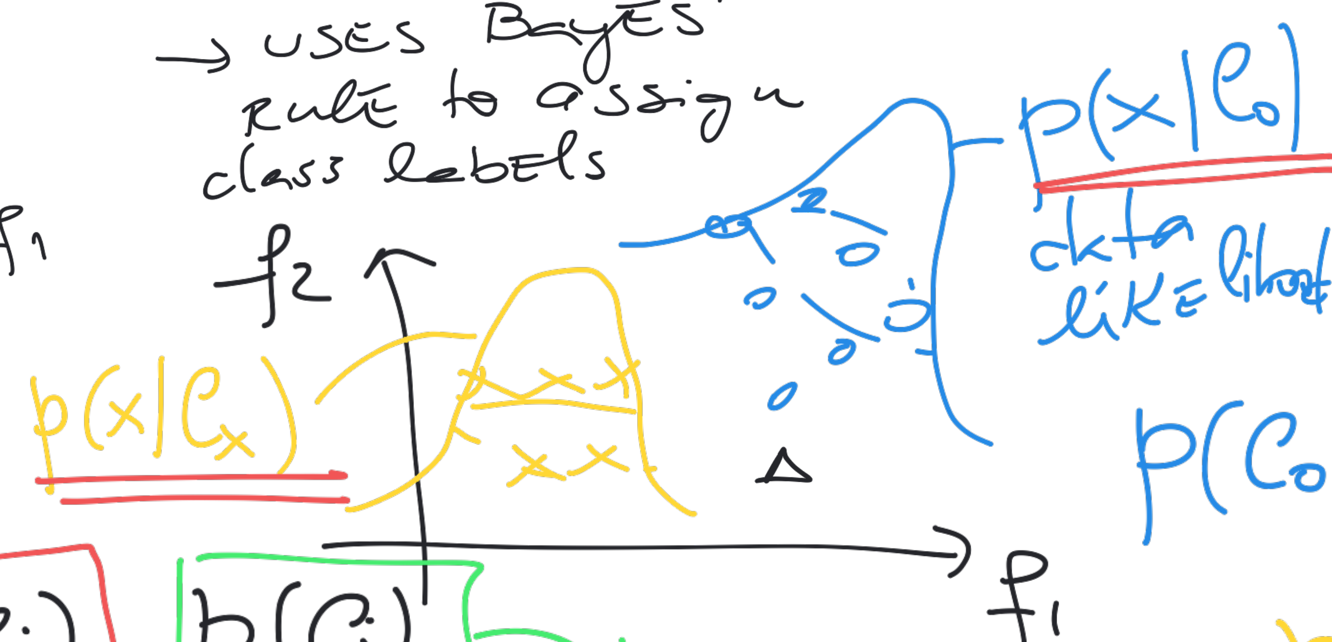
Discriminative

→ finds regions in the feature space for each class

Generative

→ finds a probabilistic model for each class

→ uses Bayes' rule to assign class labels



$$p(C_0) = \frac{7}{12}$$

$$p(C_1) = \frac{5}{12}$$

Bayes' Rule

$$\boxed{p(C_i|x)} = \frac{\boxed{p(x|C_i)} \cdot \boxed{p(C_i)}}{\boxed{p(x)}}$$

$p(C_i|x)$: posterior probability

$p(C_i)$: prior probability

$p(x)$: evidence

For every new point Δ:

$$\boxed{p(C_x|\Delta) \underset{C_o}{\overset{C_x}{\gtrless}} p(C_o|\Delta)}$$

↳ Bayesian Decision Rule

$$\Leftrightarrow \boxed{\boxed{p(\Delta|C_x)} \cdot p(C_x) \underset{C_o}{\overset{C_x}{\gtrless}} p(\Delta|C_o) \cdot p(C_o)}$$

Naïve Bayes Classifier

↳ assumes features are independent

To compute $P(x)$, use
Law of Total Probability:

$$P(x) = \underbrace{P(x \mid C_x)}\cdot\underbrace{P(C_x)} + P(x \mid C_o)\cdot P(C_o)$$

$\{C_x, C_o\}$ partitions the label sample space

Data Likelihood

→ Assume is Gaussian

So, for each $K = 1, 2$, we want to estimate the mean μ_K, and covariance Σ_K for each class's Gaussian.

$$\Sigma_K = \begin{bmatrix} \sigma_{f1}^2 & \sigma_{f_1} \cdot \sigma_{f_2} \\ \sigma_{f_2} \sigma_{f_1} & \sigma_{f_2}^2 \end{bmatrix}$$

assume Σ_K is isotropic:

$$\Sigma_K = \sigma_K^2 \cdot I = \begin{bmatrix} \sigma_K^2 & 0 \\ 0 & \sigma_K^2 \end{bmatrix}$$

$$\boxed{p(x \mid C_K) = \frac{1}{(2\pi)^{1/2} \cdot |\Sigma_K|} \cdot \exp\left\{-\frac{1}{2}(x-\mu_K)^T \Sigma_K^{-1} \cdot (x-\mu_K)\right\}}$$

- If we have $\{x_i\}_{i=1}^{N}$ (N samples), how do we compute data likelihood?

↳ assumption: samples are independent then →

$$P(x|C_k) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{1/2}|\Sigma_k|} \cdot \exp\left\{-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1} (x_i-\mu_k)\right\}$$

we want good distributional fit, i.e.

$$\boxed{\underset{\mu_k, \Sigma_k}{\arg\max}\ P(x|C_k)}$$

$$\mathcal{L} = \ln(P(x|C_k))$$

$$= \sum_{i=1}^{N} \ln\left(\frac{1}{(2\pi)^{1/2}|\Sigma_k|}\right) - \frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)$$

$$= \sum_{i=1}^{N} -\frac{1}{2}\ln(2\pi) - \ln(|\Sigma_k|) - \frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)$$

USE MLE approach to solve for μ_k and Σ_k, $\forall k = 1, 2$.

Sample mean for each class

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = 0 \iff \sum_{i=1}^{N} -(x_i-\mu_k) \cdot \underbrace{\sigma_k^2 \cdot I} = 0$$

$$\iff \sum_{i=1}^{N}(x_i-\mu_k) = 0 \iff \sum_{i=1}^{N} x_i - \sum_{i=1}^{N}\mu_k = 0$$

$$\iff \sum_i x_i - N\cdot\mu_k = 0 \iff \boxed{\mu_k = \frac{1}{N}\cdot\sum_{i=1}^{N} x_i}$$