$\boxed{\text{Ng 5}}$ Classifiers 51k views

GDA
Naive Bayes

(1)

Discriminative

$P(y|x)$ $\doteq$ learns or learn $h_\theta(x) \in \{0,1\}$

e.g. logreg: output 0, 1 directly.

Generative

$P(x|y)$

builds a probalistic model of features conditioned on class labels y.

$$P(y=1|x) = \frac{P(x|y=1)\cdot P(x)}{P(x)}$$

where $P(x) = P(y=0|x)P(x) + P(y=1|x)P(x)$

Example

Assume $x \in \mathbb{R}^n$, continuous

Gaussian discriminant analysis algorithm

Assume $P(x|y)$ is Gaussian

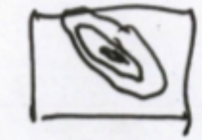
$z \sim \mathcal{N}(\vec{\mu}, \Sigma)$ ... covariance matrix $\Sigma = E[(x-\mu)(x-\mu)^T]$

(μ: mean)

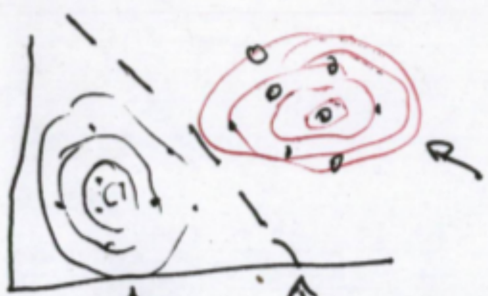
$$p(z) = \frac{1}{\sqrt{2\pi}\,|\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$ — Multivariate normal

$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & -.5 \\ -.5 & 1 \end{bmatrix}$





Useful look at Gaussians on the move.



GDA1 separator GDA2.

Benign vs Malignant cancers

$P(y) = \phi^y(1-\phi)^{1-y}$ — Bernoulli ← 'true value' of y generating data

$$P(x|y=0) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(x-\mu_0)^T \Sigma^{-1}(x-\mu_0)\right)$$

$$P(x|y=1) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)\right)$$

$$\ell(\phi, \mu_0, \mu_1, \Sigma) = \log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)})$$ ← joint likelihood

$$= \log \prod_i p(x^{(i)}|y^{(i)}) \cdot p(y^{(i)})$$

cf logreg

$$\ell(\theta) = \log \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta)$$ ← conditional likelihood

To fit, do $\boxed{\text{MLE}}$

$\max \ell(\phi, \mu_0, \mu_1, \Sigma)$

$$\phi_{MLE} = \frac{\sum_i y^{(i)}}{m} = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=1\}}{m}$$: fraction with label '1'

$$\mu_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=0\}\,x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}}$$: examples with label '0'