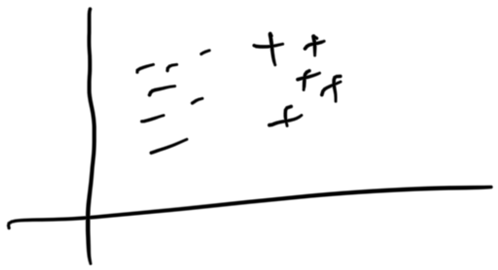
COL774
Machine Learning
Sep 7, 2021

Last Class:-

GDA: Gaussian Discriminant Analysis

$y^{(i)}|x^{(i)}; \theta \sim$



↳ $x^{(i)}, y^{(i)}; \theta$

↳ $y^{(i)}; \theta \sim \text{Bernoulli}(\phi)$ ①

$x^{(i)}|y^{(i)}; \theta \sim N(\mu, \Sigma)$

$$\theta \equiv (\phi, \mu_0, \Sigma_0, \mu_1, \Sigma_1)$$

$$x \sim N(\mu, \sigma^2), \quad x \in \mathbb{R}$$

$$p(x) = \frac{1}{\sqrt{2\pi}(\sigma^2)^{1/2}} \cdot e^{-\frac{(x-\mu)(\sigma^2)^{-1}(x-\mu)}{2}}$$

$\Downarrow$

$x \in \mathbb{R}^n$

$$p(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}}$$

$\arg\max_{\theta} L(\theta) \to$

$$\phi = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=1\}}{m}$$

$$\mu_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=0\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}} \qquad \mu_1 = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=1\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)}=1\}}$$

$$\hat{\Sigma}_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=0\} (x^{(i)}-\mu_0)(x^{(i)}-\mu_0)^T}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}} \qquad \Sigma_1 = \sim \text{Similar}$$

$x^i \quad y^{(i)}=0$

$\quad\;\; y^{(i)}=1$

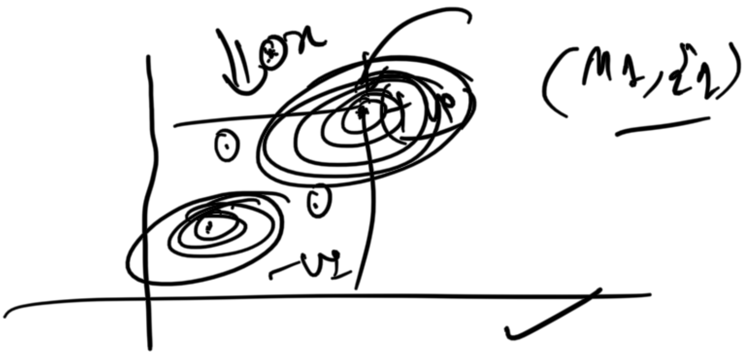
("tied")

if $\hat{\Sigma}_0 = \hat{\Sigma}_1 = \hat{\Sigma}$

$$\hat{\Sigma} = \frac{\sum_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T}{m}$$

try this $n=1$

$$\theta = (\phi, \mu_0, \hat{\Sigma}_0, \mu_1, \hat{\Sigma}_1) \; ; T$$



$P(y;\theta)$ ✓

$P(x|y;\theta)$ ✓

$$P(y|x;\theta) = ?$$

Use Bayes Rule

$$= \frac{P(x|y;\theta)\,P(y;\theta)}{P(x)}$$

$$P(y=1|x;\theta) = \frac{P(x|y=1;\theta)\,P(y=1;\theta)}{P(x)}$$

$P(x) = ?$ $\quad P(x) = P(x|y=0;\theta)\,P(y=0;\theta) + P(x|y=1;\theta)\,P(y=1;\theta)$

$$\sum_y P(x|y)\,P(y)$$

$$P(y=1|x;\theta) = \frac{P(x|y=1;\theta)\,P(y=1;\theta)}{P(x|y=1;\theta)\,P(y=1;\theta) + P(x|y=0;\theta)\,P(y=0;\theta)}$$

$$= \frac{1}{1 + \left[\frac{P(x|y=0;\theta)\,P(y=0;\theta)}{P(x|y=1;\theta)\,P(y=1;\theta)}\right]}$$

A

$$= \frac{1}{1+A}$$

Log A

$\Rightarrow$ Decision Boundary

$P(y=1|x;\theta)$

$= 1/2$

$\Rightarrow A = 1 \Rightarrow \log A = 0$

Case 1:- $\Sigma_0 \neq \Sigma_1$

Case 2:- $\Sigma_0 = \Sigma_1$

$$A = \frac{P(x|y=0;\theta)\quad P(y=0;\theta)}{P(x|y=1;\theta)\quad P(y=1;\theta)}$$

$$= \frac{1}{\cancel{(2\pi)^{n/2}}|\Sigma_0|^{1/2}}\, e^{-\frac{(x-\mu_0)^T\Sigma_0^{-1}(x-\mu_0)}{2}} \cdot (1-\phi)$$

$$\frac{1}{\cancel{(2\pi)^{n/2}}|\Sigma_1|^{1/2}} e^{-\frac{(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)}{2}} \cdot(\phi)$$

$$\log A = \log\left[\left[\frac{1-\phi}{\phi}\right]\cdot\frac{|\Sigma_1|^{1/2}}{|\Sigma_0|^{1/2}}\right] \equiv C$$

$$-\frac{1}{2}\left[(x-\mu_0)^T\Sigma_0^{-1}(x-\mu_0) - (x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right]$$

$$= \frac{1}{2}\left[(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - (x-\mu_0)^T\Sigma_0^{-1}(x-\mu_0)\right] + C$$

$$-\mu_1^T\Sigma_1^{-1}x$$
$$-x^T\Sigma_1^{-1}\mu_1$$
$$-\mu_1^T\Sigma_1^{-1}x$$

$$= \frac{1}{2}\left[x^T\Sigma_1^{-1}x - \underbrace{2\mu_1^T\Sigma_1^{-1}x} + \mu_1^T\Sigma_1^{-1}\mu_1 - \left(x^T\Sigma_0^{-1}x - 2\mu_0^T\Sigma_0^{-1}x + \mu_0^T\Sigma_0^{-1}\mu_0\right]\right.$$

$$= \frac{1}{2}\left[\overbrace{x^T(\Sigma_1^{-1}-\Sigma_0^{-1})x}^{\text{Quadratic}} \overbrace{-2(\mu_1^T\Sigma_1^{-1}-\mu_0^T\Sigma_0^{-1})x}^{\text{Linear}} + \mu_1^T\Sigma_1^{-1}\mu_1 - \mu_0^T\Sigma_0^{-1}\mu_0\right] + C \equiv \log A$$

$+ C$

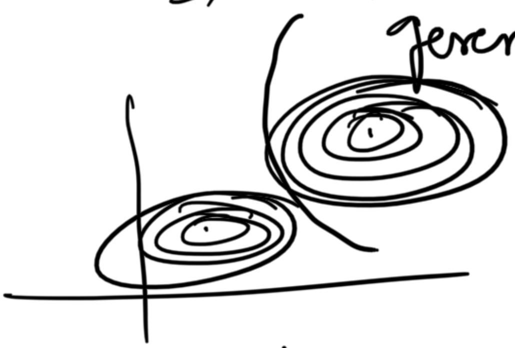
∴ Above is a quadratic expression

In general:- ~~Assum~~ is ... in x

$\Rightarrow$ At decision Boundary,

$\log A = 0$

$\Rightarrow$ & the separator is quadratic in general for GDA based model



Case 2:- $\Sigma_0 = \Sigma_1 = \Sigma$

$$\log A = -\frac{1}{2}\Big[2(\mu_1 - \mu_0)^T \Sigma^{-1} x - \mu_1^T \Sigma^{-1} \mu_1 + \mu_0^T \Sigma^{-1} \mu_0\Big] + C$$

$$= -\Big[(\mu_1 - \mu_0)^T \Sigma^{-1} x - \frac{1}{2}\big[\mu_1^T \Sigma^{-1} \mu_1 + \mu_0^T \Sigma^{-1} \mu_0\big] - C\Big]$$

$\Rightarrow$ Decision Boundary is linear

$\log A = -\theta^T x$

GDA

$$p(y=1 \mid x; \theta) = \frac{1}{1+A} = \frac{1}{1+e^{(\log A)}}$$

$\Sigma = 0$

⇒ under the assumption of identical co-variance matrices

↳ ① Decision Boundary is linear

of form of $p(y=1 \mid x; \theta) = \frac{1}{1+e^{-\theta^T x}}$

logistic

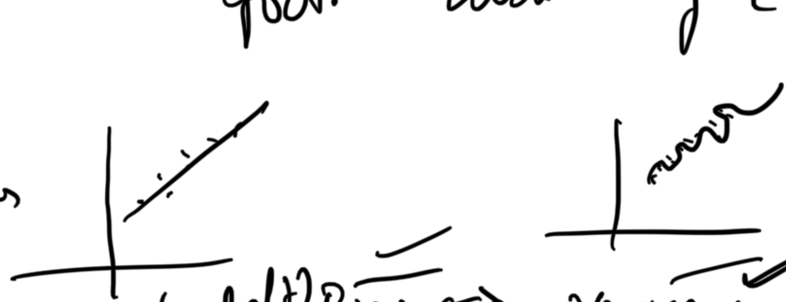
Discriminative vs Generative

$\Sigma_0 = \Sigma_1 = \Sigma$ → GDA ⇒ logistic decision Boundary

(I) which model makes stronger assumptions about the data?] GDA

(II) which model is more/less likely to overfit?]

(III) for a real problem, which will be your model of choice?



(underfit) Bias ⟷ variance (overfit)

Trade-off

Naive Bayes :→ Generative Model

↳ Independence of attributes | class

$x_j \perp x_k \mid y \quad \forall j,k$

$x_1 \ldots x_n \mid y \quad \because$