Generative models

→ Discriminative → $p(y|x)$

$x \longrightarrow y$

$x \to M \to y$

→ Gen: $p(x|y)$ and $p(y)$

$c = 0 \leftarrow$ cats
$c = 1 \leftarrow$ dogs

$p(x|y=0)$ features of cats

$p(x|y=1)$ features dogs

$p(y) \longrightarrow$ Bernoulli

0.6 0.4

Classifier → $p(\underset{?}{y}|\underline{x})$

$p(x|y)$ & $p(y) \longrightarrow p(y|x)$

Bayes Theorem

$$p(y|x) = \frac{p(x|y) \cdot p(y)}{p(x)} \leftarrow ? \longleftarrow ①$$

$$p(x) = p(x|y=0) \cdot p(y=0) + p(x|y=1) \cdot p(y=1)$$

$$= \sum_c p(x|y=c)\, p(y=c)$$

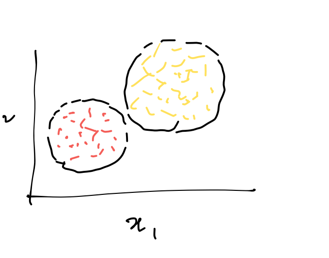
(X, Y)

$\underbrace{p(x|y)}$ and $p(y)$

→ Gaussian

GDA: Gaussian Discriminant Analysis

$x = \mathbb{R}^2 \quad x \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \leftarrow$



$p(x|y=r) = N(\mu_0, \Sigma_0) \leftarrow$

$p(x|y=y) = N(\mu_1, \Sigma_1) \leftarrow$

↑ Gaussian

$p(y) = \text{Bernoulli}(\phi) \leftarrow$

$0.6 \to 1$
$0.4 \to 0$

$$\hat{y} = \underset{y}{\arg\max}\; p(y|x)$$

$$= \underset{y}{\arg\max}\; \underline{p(x|y)\, p(y)}$$

$$\hat{y} = \underset{y}{\arg\max}\; p(x|y)\, p(y)$$

$p(x|y=0) = N(\mu_0, \Sigma_0)$

$p(x|y=1) = N(\mu_1, \Sigma_1)$

$p(y) = \text{Bern}(\phi)$

$$\text{Likelihood} = \sum_{i=1}^{m} p(x, y; \underline{\mu_0, \mu_1, \Sigma_1, \Sigma_1, \phi})$$

$$= \sum_{i=1}^{m} p(x|y; \mu_0, \mu_1, \Sigma_0, \Sigma_1)\, p(y; \phi)$$

estimate $\mu_0, \mu_1, \Sigma_0, \Sigma_1, \phi$

Training data

Mean, Cov.

$$\mu_0 = \frac{\sum_{i=1}^{m} x^{(i)}\, 1\{y^{(i)}=0\}}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}}$$ average of datapoints belonging to $y=0$.

$$1\{y^{(i)}=0\} = f(y) = \begin{cases} 1, & y=0 \\ 0, & y=1 \end{cases}$$

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

$$\Sigma_0 = \frac{1}{1\{y=0\}} \sum_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T \leftarrow$$

$\mu_{y^{(i)}} \to \mu_0$
μ_1

$$\Sigma_1 = \frac{1}{1\{y=1\}} \sum_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$$

$$\phi = p(y=1) = \frac{\text{favourable outcomes}}{\text{all outcomes}}$$

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

Parameters $= (\mu_0, \mu_1, \Sigma_0, \Sigma_1, \phi) \leftarrow$
↳ maximize L.

How to classify

$p(y=0|x)$ and $p(y=1|x)$

$\underbrace{p(x|y=0)}_{N(\mu_0, \Sigma_0)}\; \underbrace{p(y=0)}_{\text{Ber}(1-\phi)}$ $\quad \underbrace{p(x|y=1)}_{N(\mu_1, \Sigma)}\; \underbrace{p(y=1)}_{\text{Bern}(\phi)}$

$$p(x|y=0) = \frac{1}{(2\pi)^{d/2} |\Sigma_0|^{1/2}} \exp\left\{ \frac{-1}{2} (x-\mu_0)^T \Sigma_0^{-1} (x-\mu_0) \right\}$$

$\underbrace{\Sigma_0 \;\; \Sigma_p \;\; \Sigma_2}$
$\to \Sigma \to 1000$
$10 \to 1000 \times 1000$

$p(y=0) = 1 - \phi$

$\text{Num}_0 \to P_0$

Similarly → $\text{Num}_1 \to P_1$

GDA → classifier

Revise

→ $p(y|x) \leftarrow$

→ $p(x|y)$ and $p(y)$

→ Bayes Theorem
$p(x|y)$ & $p(y) \longrightarrow p(y|x)$

→ Gaussian → GDA
$(\mu, \Sigma)\; \phi$
Training $\mu_0, \Sigma_0, \phi,$

→ $p(y|x=c) = p(x|y=c)\, p(y=c)$

2 > >

Multi-class

$y = 0 \to y = 1$

Bernoulli

$$p(y) = \begin{cases} 1 & \phi \\ 0 & 1-\phi \end{cases}$$

$\phi = 0.5$

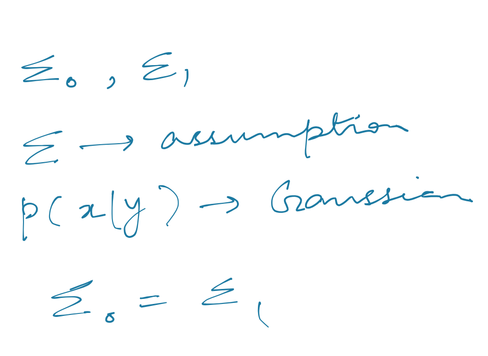
Multinoulli
↳ Dice
$\phi_0 \quad \phi_1 \quad \phi_K$
$\sum \phi = 1$

$$\phi_0 = \frac{1\{y^{(i)}=0\}}{m}$$

$$\phi_1 = \frac{1\{y^{(i)}=1\}}{m}$$

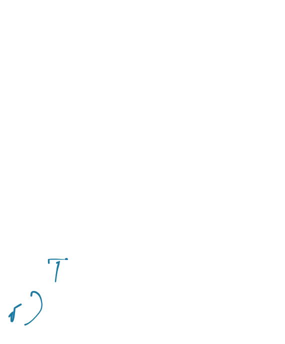
$y=0 \longrightarrow \mu_0, \Sigma_0$
$y=1 \longrightarrow \mu_1, \Sigma_1$

$x = \begin{bmatrix} \vdots \end{bmatrix}$



→ μ_0

→ Σ_0



Σ_0, Σ_1

$\Sigma \longrightarrow$ assumption

$p(x|y) \to$ Gaussian

$\Sigma_0 = \Sigma_1$

$$\Sigma_0 = \sum_{i}^{m} (x_i - \mu_0)(x_i - \mu_0)^T$$

Σ_1

$\underbrace{\Sigma_0 \quad \Sigma_1}_{\text{Powerful}} \qquad \Sigma$

Decision boundary

$\underbrace{p(y|x=0) = p(y|x=1)}$

Shared $\Sigma \longrightarrow$ Linear $\longleftarrow$ LDA

Individual Σ's $\longrightarrow$ Quadratic $\longleftarrow$ QDA