Bayesian Classifiers.

$P(Y^* = \text{benign} \mid x^*)$

$P(Y^* = \text{malignant} \mid x^*)$

x_1	y_1	M	— H
x_2	y_2	M	— H
x_3	y_3	B	— T
x_4	y_4	M	— H
x_5	y_5	B	— T

$P(Y^* = \text{benign}) = \frac{2}{5}$

$P(Y^* = \text{malignant}) \longrightarrow = \frac{3}{5}$

$\boxed{Y^* \text{ is a Bernoulli random variable}}$

$$P(Y^* = \text{malignant} \mid X^* = x^*)$$

$$= \frac{P(X^* = x^* \mid Y = \text{malignant})\, P(Y = \text{malignant})}{P(X = x^* \mid Y = \text{malignant})\, P(Y = \text{malignant}) + P(X = x^* \mid Y = \text{benign})\, P(Y = \text{benign})}$$

$P(\underline{X = x^*} \mid Y = \text{malignant})$

	Y
x_1	M
x_2	M
x_3	M
x_4	B
x_5	B
x_6	M
x_7	B
x_8	M

x^*

$Y = \text{mal.}$ $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_6 \\ x_8 \end{bmatrix}$ $Y = \text{ben.}$ $\begin{bmatrix} x_4 \\ x_5 \\ x_7 \end{bmatrix}$

~~Agg~~ Assignment 3 - Released Tonight - May 8th

Office Hours - 2-3 PM

D = 5

$X = (x_1, x_2, x_3, x_4, x_5)$

binary.

X can take 2^5 values.

Joint probability Distribution

Let $x_1, x_2, x_3, x_4, x_5 \in \{0, 1\}$

0	0	0	0	0
0	0	0	0	1
1	0	0	0	0
⋮				

2^5

We need 2^5 probabilities in the joint probability distribution.

Technically we need $(2^5 - 1)$ probabilities.

Let D = 100 ; $2^{100} - 1$ probabilities need to be estimated

$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)\, p(x_2)\, p(x_3)\, p(x_4)\, p(x_5)$

Assuming that x_1, x_2, x_3, x_4, x_5 are independent of each other

# of probabilities that we need: $1 + 1 + 1 + 1 + 1 = 5$

$$\boxed{2^D - 1 \Rightarrow D}$$

let $Y \in \{1, 2\}$ $\quad X_j \in \{1, 2\}$

let $D = 5$

~~$P(X = x^*$~~

$$P(Y=1 \mid X = \{x_1, x_2, x_3, x_4, x_5\}) = \frac{P(X = \{x_1, x_2, x_3, x_4, x_5\} \mid Y=1) * P(Y=1)}{(\qquad)}$$

$$= \frac{P(X_1 = x_1 \mid Y=1)\, P(X_2 = x_2 \mid Y=1) \cdots\cdots P(Y=1)}{(\qquad)}$$

If $x_1 = 1$, $x_2 = 1$, $x_3 = 1$, $x_4 = 1$, $x_5 = 1$

$$= \frac{\theta_{11}\,\theta_{12}\,\theta_{13}\,\theta_{14}\,\theta_{15}\,\theta_1}{(\qquad)}$$

$$= \frac{\theta_1 \prod_{j=1}^{D} \theta_{1j}}{\left(\theta_1 \prod_{j=1}^{D} \theta_{1j} + \theta_2 \prod_{j=1}^{D} \theta_{2j}\right)}$$

$$\boxed{\text{Training time} \rightarrow \text{We need to estimate } \theta_1, (\theta_{1j}) \quad \theta_2, (\theta_{2j})}$$

What is θ_1? Assuming Y is Bernoulli:

$\hat{\theta}_1 = \frac{N_1}{N}$ — # of examples where $Y=1$

θ_2

$\hat{\theta}_2 = \frac{N_2}{N}$ — # of examples where $Y=2$

$$\theta_{1j} = \frac{N_{1j}}{N_1} \quad \begin{array}{l} \text{— \# feature } j \text{ take value 1 when } Y=1 \\ \longrightarrow \text{\# examples where } Y=1 \end{array}$$

$$\theta_{2j} = \frac{N_{2j}}{N_2} \quad \begin{array}{l} \text{— \# feature } j \text{ take value 1 when } Y=2 \\ \longrightarrow \text{\# examples where } Y=2 \end{array}$$

These are all <u>MLE</u>

<u>Adding a prior:</u>

<u>To Y</u> $\quad Beta(a, b)$

$$\hat{\theta}_1 = \frac{N_1 + a}{N + a + b}$$

To <u>$X_j | Y$</u> $\quad Beta(a_j, b_j)$

$$\hat{\theta}_{1j} = \frac{N_{1j} + a_j}{N_1 + a_j + b_j}$$

$$\hat{\theta}_{2j} = \frac{N_{2j} + a_j}{N_2 + a_j + b_j}$$

Data:

Shape	Size	Color	Type
Cir	lg	li	m
cir	lg	li	b
cir	lg	li	m
ov	lg	li	b
ov	lg	dk	m
ov	sm	dk	b
ov	sm	dk	m
ov	sm	li	b
cir	sm	dk	b
Cir	lg	dk	m
cir-1	lg-1	li-1	m-1
ov-2	sm-2	dk-2	b-2

$X^* = ci, sm, li$

$\theta_1 (\theta_m) = \frac{5}{10} = \frac{1}{2}$ $\theta_2(\theta_b) = \frac{5}{10} = \frac{1}{2}$

$\theta_{1\,shape} = \frac{3}{5}$	$\theta_{1\,size} = \frac{4}{5}$	$\theta_{1\,color} = \frac{2}{5}$
$\theta_{2\,shape} = \frac{2}{5}$	$\theta_{2\,size} = \frac{2}{5}$	$\theta_{2\,color} = \frac{3}{5}$

$p(y = \text{malignant} \mid X = \underset{1}{cir}, \underset{2}{sm}, \underset{1}{li})$

$$= \frac{\cancel{\left(\frac{5}{10}\right) *}\ \frac{1}{2} * \frac{3}{5} * \left(1 - \frac{4}{5}\right) * \frac{2}{5}}{\left(\frac{1}{2} * \frac{3}{5} * \frac{1}{5} * \frac{2}{5}\right) + \left(\frac{1}{2} * \frac{2}{5} * \left(1 - \frac{2}{5}\right) * \frac{3}{5}\right)} =$$

~~Si~~ (Size | Y = M) - Bernoulli random variable

Size = lg

= sm

$\theta_{1\,size} \rightarrow$ Prob. of (Size) to take value lg when class is malignant. (Y = 1)

$$P(\text{Size} = \text{sm} \mid Y = \text{malignant}) = (1 - \theta_{1\,size})$$

Building a NBC from text:

X	y
~~a~~ great game	Sports
~~the~~ election ~~was~~ over	Non-Sports
Very clean ~~and~~ close match	Sports
A clean ~~but~~ forgettable game	Sports
~~It was a~~ close election	non-Sports.

→ Stop-word removal

→ Stemming. (x)

close, closed → close

→ Term vector representation

→ vocabulary building.

great game election over very clean match forgettable close

great	game	election	over	very	clean	match	forgettable	close	tag
1	1	0	0	0	0	0	0	0	1
0	0	1	1	0	0	0	0	0	0
0	0	0	0	1	1	1	0	1	1
0	1	0	0	0	1	0	1	0	1
0	0	1	0	0	0	0	0	1	0

~~P(tag=~~

Test: " ~~A very~~ close game" ?

x^* | 0 1 0 0 0 0 0 0 1 |

$$P(Tag=1 \mid x^*) = \left(\frac{3}{5}\right) * \left(\frac{2}{3} * \frac{2}{3} * 1 * 1 * \frac{1}{3} * \frac{1}{3} * \frac{2}{3} * \frac{2}{3} * \frac{1}{3}\right)$$

$$= a \qquad \left(\frac{a}{a+b}\right)$$

$$P(Tag=0 \mid x^*) = b \qquad \left(\frac{b}{a+b}\right)$$

Bag of Words Representation (BOW)

What if x_j is continuous? $1 \le j \le D$

$$P(y=1 \mid x) = \frac{P(y=1)\prod_{j=1}^{D} P(x_j \mid y=1)}{\text{~~Norm~~ (Normalization Constant)}}$$

What is $P(x_j \mid y=1)$?

We will assume that $(x_j \mid y=1) \sim \mathcal{N}(x_j \mid \mu_j, \sigma_j^2)$

$$P(y=1 \mid \mathbf{x}) = P(y=1) \prod_{j=1}^{D} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right]$$

$$= P(y=1) \frac{1}{(2\pi)^{D/2}(\sigma_1^2\sigma_2^2\cdots\sigma_D^2)^{1/2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{D}\frac{(x_j-\mu_j)^2}{\sigma_j^2}\right]$$

Consider a matrix (D×D) $\Sigma = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \sigma_3^2 \ddots & \\ 0 & & & \sigma_D^2 \end{bmatrix}$

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & 0 \\ 0 & \frac{1}{\sigma_2^2} & \\ & & \ddots \frac{1}{\sigma_D^2} \end{bmatrix}$$

$$|\Sigma| = \sigma_1^2\sigma_2^2\cdots\sigma_D^2$$

let $\underset{D\times 1}{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$ $\underset{D\times 1}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_D \end{bmatrix}$

$$\sum_{j=1}^{D} \frac{(x_j-\mu_j)^2}{\sigma_j^2} \equiv (\mathbf{x}-\mu)^T \Sigma^{-1} (x-\mu)$$

If $\Sigma_1 = \Sigma_2 = \Sigma$ then QDA becomes LDA

linear

Ignoring $P(y=1)$ $P(y=0)$ and let $\Sigma = \Sigma_1 = \Sigma_2$

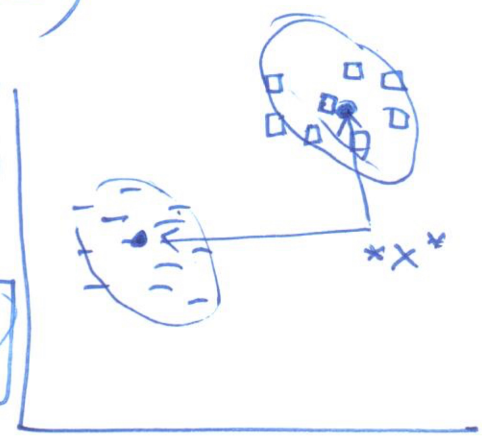
LDA

$$P(y=1 \mid x^*) \propto \mathcal{N}(x^* \mid \mu_1, \Sigma)$$

$$P(y=0 \mid x^*) \propto \mathcal{N}(x^* \mid \mu_2, \Sigma)$$

$$\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)\right]$$

$$\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-(x-\mu_2)^T \Sigma^{-1} (x-\mu_2)\right]$$



$(x-\mu_1)^T \Sigma^{-1} (x-\mu_1) \rightarrow$ Mahalanobis distance between x & μ_1

$$\boxed{(x-\mu_1)^T (x-\mu_1) \rightarrow \text{Euclidean distance}}$$

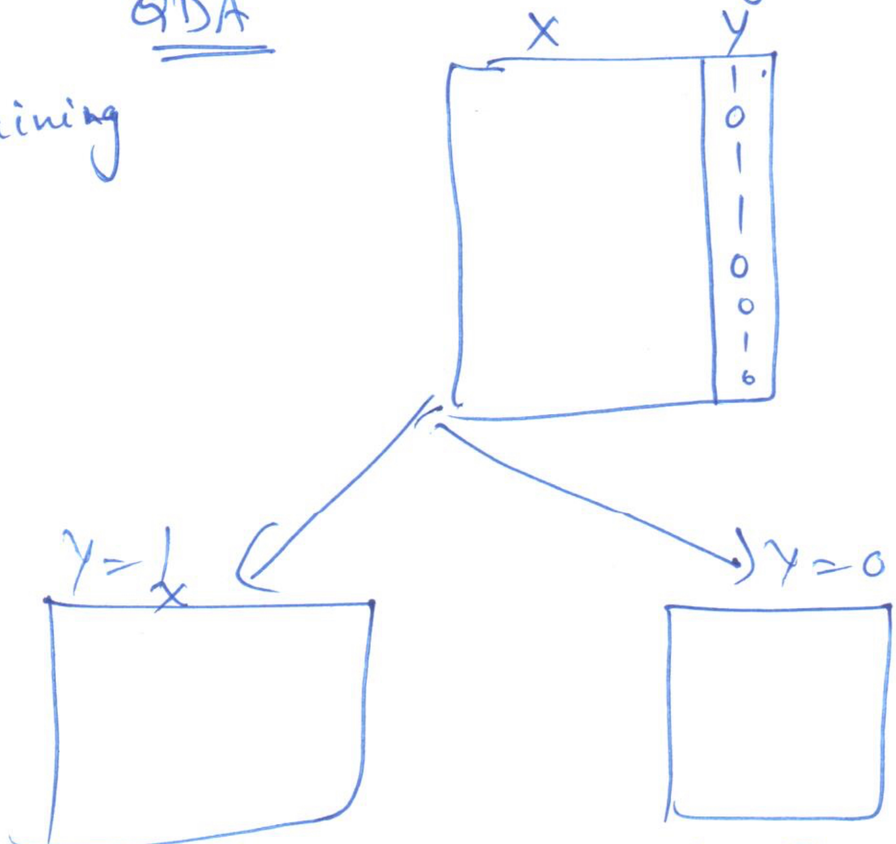
$$P(Y=1 \mid x) = P(Y=1)\,\underbrace{\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]}_{\text{pdf of a multivariate Gaussian } (\mu,\Sigma)}$$

$$\boxed{\text{Same as assuming that } x \sim N(\mu, \Sigma)}$$

Quadratic Discriminant Analysis Classifier

<u>QDA</u>

Training



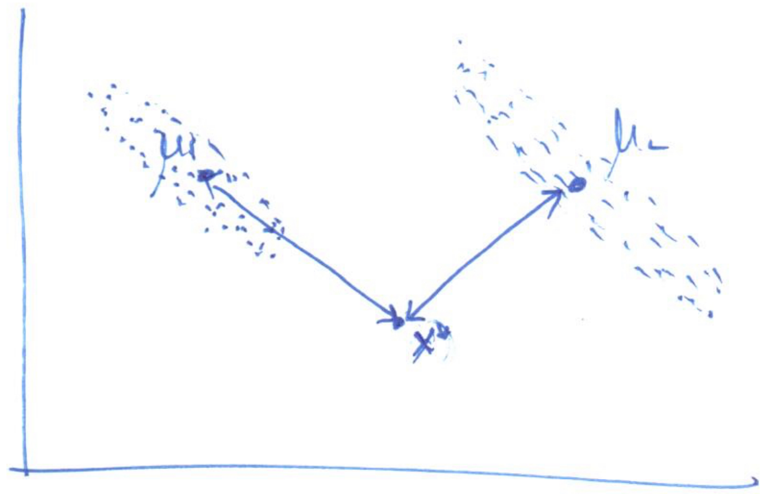
estimate $\mu_1\ \Sigma_1$ (MLE) $\qquad \mu_2\ \Sigma_2$ (MLE)

Testing

x^*

$$P(Y=1 \mid x^*) = \frac{P(Y=1)\,N(x^* \mid \mu_1, \Sigma_1)}{P(Y=1)\,N(x^* \mid \mu_1, \Sigma_1) + P(Y=0)\,N(x^* \mid \mu_2, \Sigma_2)}$$

Mahalanobis Distance



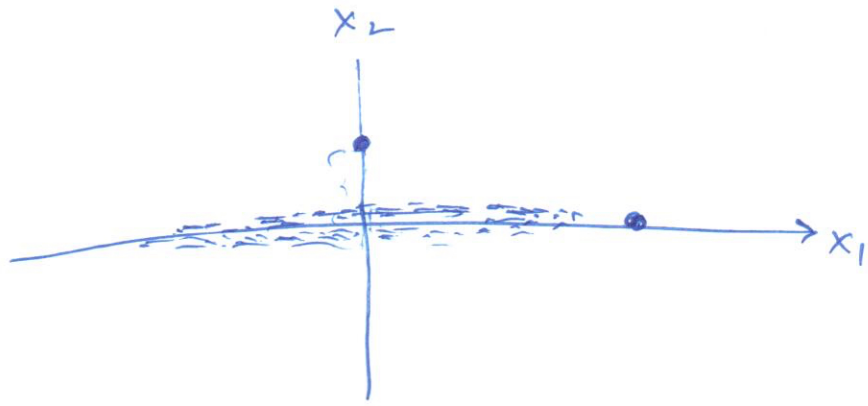
$$\frac{\|x^* - \mu_1\|_2}{\|x^* - \mu_2\|_2}$$

$$(x^* - \mu_1)^T (x^* - \mu_1)$$

$$(x^* - \mu_2)^T (x^* - \mu_2)$$

let us assume that $\underline{(x^* - \mu_1)^T (x^* - \mu_1)} = (x^* - \mu_2)^T (x^* - \mu_1)$

$$\underline{(x^* - \mu_1)^T \Sigma_1^{-1} (x^* - \mu_1)} \ll \underline{(x^* - \mu_2)^T \Sigma_2^{-1} (x^* - \mu_2)}$$



QDA	LDA
$\Sigma_1 \neq \Sigma_2$	$\Sigma_1 = \Sigma_2 = \Sigma$
$\frac{1}{(2\pi)^{D/2} \lvert\Sigma_1\rvert^{1/2}} \exp\left[-\frac{1}{2} (x-\mu_1)^T \Sigma_1^{-1} (x-\mu_1)\right]$	$\frac{1}{(2\pi)^{D/2} \lvert\Sigma\rvert^{1/2}} \exp\left[-\frac{1}{2} (x-\mu_1)^T \Sigma^{-1} (x-\mu_2)\right]$
$\frac{1}{(2\pi)^{D/2} \lvert\Sigma_1\rvert^{1/2}} \exp\left[-\frac{1}{2} (x-\mu_2) \Sigma_2^{-1} (x-\mu_2)\right]$	$\frac{1}{(2\pi)^{D/2} \lvert\Sigma\rvert^{1/2}} \exp\left[-\frac{1}{2} (x-\mu_1)^T \Sigma^{-1} (x-\mu_2)\right]$