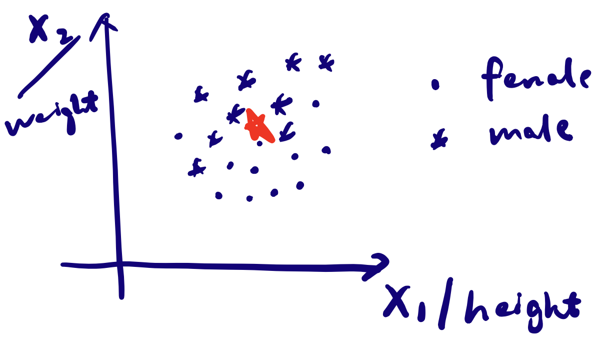
Sep 10, 2024

$P(\underline{x} | y = \text{malignant}) = P(x_1 | y = m) \cdot P(x_2 | y = m) \cdot P(x_3 | y = m)$

shape / size / color $\quad \sim Ber(\theta_1) \quad \sim Ber(\theta_2) \quad \sim Ber(\theta_3)$



$P(y | x^*) \propto P(y) \cdot \underline{P(x^* | y)}$

$P(x^* | y) = \underset{\sim G(\mu_1, \sigma_1)}{P(x_1 | y)} \cdot \underset{\sim G(\mu_2, \sigma_2)}{P(x_2 | y)}$ Naive Assumption

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_D^2 \end{bmatrix} \begin{cases} 0 \text{ covariance} \\ \text{independent attributes} \end{cases}$$

Gaussian Discriminant Analysis

Given $\begin{cases} x = [x_1, x_2] & \text{Continuous} \\ y : 0/1 & \text{Binary} \end{cases}$

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

$P(x | y = 0) \sim N(\mu_2, \Sigma_2)$

$P(y | x^*) \propto \underline{P(x^* | y) \cdot P(y)}$

$\sim N(\mu_1, \Sigma_1) \cdot \sim Ber(\theta)$

$N(\mu_2, \Sigma_1)$

Estimate $\mu_1, \Sigma_1, \mu_2, \Sigma_2, \phi$ from traing data.

Split data based on y 0/1

$$\left\{\begin{array}{l} x_1 \\ \vdots \\ x_m \end{array}\right. \quad y=1 \quad \xRightarrow{MLE} \quad \mu_1\ \Sigma_1$$

$$\left\{\begin{array}{l} x_1 \\ \vdots \\ x_p \end{array}\right. \quad y=0 \quad \Rightarrow MLE \quad \mu_2\ \Sigma_2$$

$$\phi_{MLE} = \frac{m}{m+p}$$

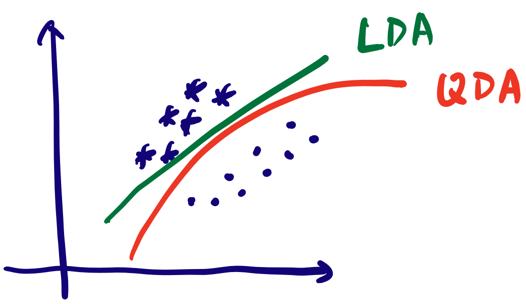
testing : $P(y|x^*)$ QDA

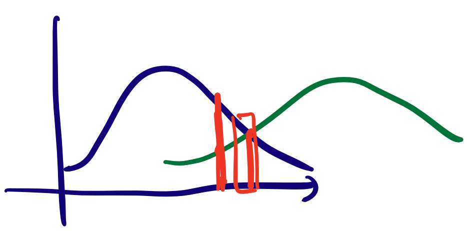
$\Sigma_1 = \Sigma_1 = \Sigma$. assumption $\Rightarrow$ LDA

Linear Discriminant Analysis

training μ_1, μ_2 ϕ the same as QDA

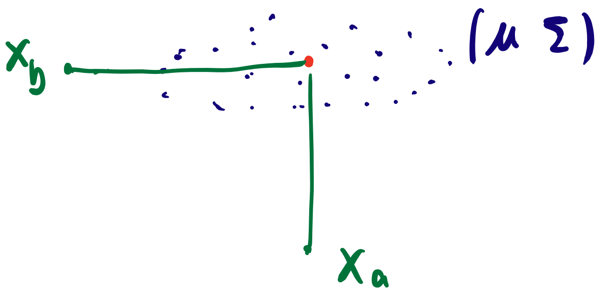
Σ use all traing data





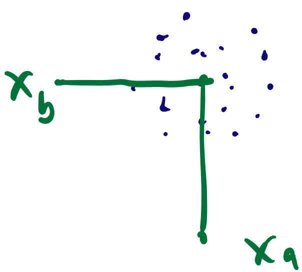
Mahalanobis Distance

$$(x-\mu)^T \Sigma^{-1} (x-\mu)$$

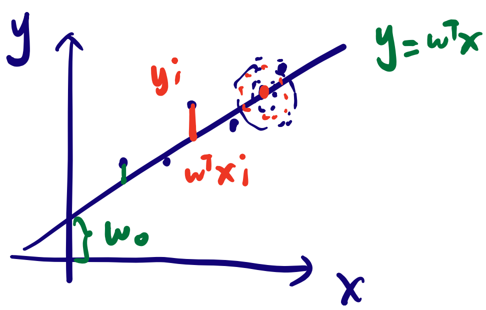


Euclidean Distance

$(x-\mu)^T (x-\mu)$, $\Sigma = I$



Linear regression



Probabilistic interpretation

$y \sim N(w^T x, \sigma^2)$

$y = w^T x + \varepsilon$

$\varepsilon \sim N(0, \sigma^2)$

training: estimate w, σ

Geometric interpretation, line fitting

$y = w^T x$ $\qquad$ $y = w^T \underset{0}{\underset{\uparrow}{x}} = 0$

$$y = w_0 + w^T x$$

bias, intercept

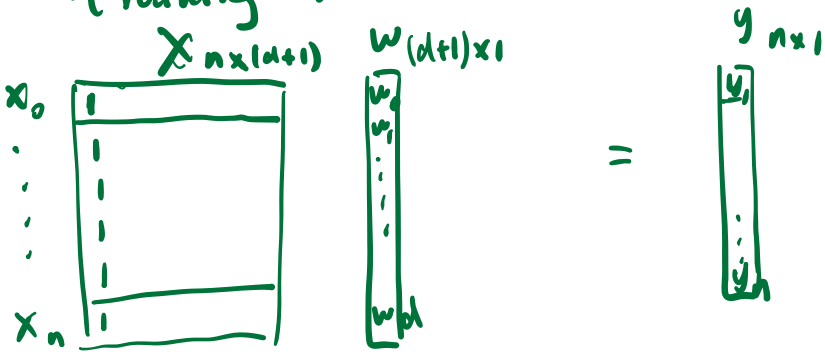
augment $\quad x \to [1, x] \quad [1, \text{height}, \text{weight}]$

$\quad\quad\quad\quad w \to [w_0, w]$

$$y = \underset{[w_0,\, w]}{w^T}\ \underset{[1,\, x]}{x}$$

$$= w_0 + w_1 x_1 + w_2 x_2 + \cdots w_d x_d$$

training data



$$X \cdot w = y$$

Likelihood function

$$L = \prod_{i=1}^{N} \underline{\mathcal{N}(w^T x, \sigma^2)}$$

$$\log L = \sum_{i=1}^{N} \left(-\log\sqrt{2\pi} - \log\sigma - \underline{\frac{1}{2\sigma^2}(y_i - w^T x_i)^2}\right)$$

to estimate w, set $\dfrac{\partial \log L}{\partial w} = 0$

$$\frac{\partial \log L}{\partial w} = -\frac{1}{2\sigma^2}\frac{\partial}{\partial w}(y_i - w^T x_i)^2 = 0$$

$$\boxed{w = (X^T X)^{-1} X^T y}$$

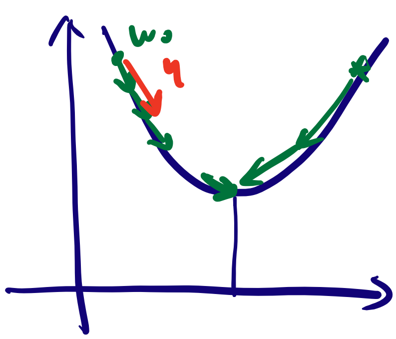
to estimate σ set $\frac{\partial \log L}{\partial \sigma} = 0$

Geometric interpretation

$$\min \sum_{i=1}^{N} \varepsilon$$

$$\min \sum_{i=1}^{N} (y_i - w^T x_i)^2 = J(w)$$

$$\frac{\partial J(w)}{\partial w} = 0 \qquad w =$$



Gradient Descent

Initialize w_0

$$w_{i+1} = w_i - \eta \frac{\partial J(w_i)}{\partial w_i}$$

learning rate