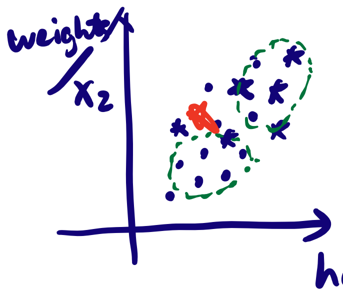
09/14/2022 Wed

$$P(x \mid y = \text{malignant}) = p(\text{shape} \mid y=m)\,p(\text{size} \mid y=m)\,p(\text{color} \mid y)$$

x shape | size | color $\sim Ber(\theta_1)$ $\sim Ber(\theta_2)$ $\sim Ber(\theta_3)$



• female

* male

$$P(y=1 \mid x) = \frac{P(x \mid y=1)\,P(y=1)}{\sum_y (x \mid y)\,P(y)}$$

$$P(x \mid y=1) = P(\text{height} \mid y=1)\,P(\text{weight} \mid y=1)$$ ← Naive Bayes Assumpsion

$\sim G(\mu_1, \sigma_1)$ $\sim G(\mu_2, G_2)$

$$P(y \mid x) \propto P(y) \cdot \prod_{j=1}^{D} P(x_j \mid y)$$

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

let $\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots & \sigma_D^2 \end{bmatrix}$ $\quad \prod_{j=1}^{D} \sigma_j = |\Sigma|^{\frac{1}{2}}$

let $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_j \end{bmatrix}$

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

Given $X = [x_1 \ x_2]$

y

Gaussian Discriminant Analysis

$$P(x|y=1) = \mathcal{N}(\mu_1 \ \Sigma_1)$$
$$P(x|y=2) = \mathcal{N}(\mu_2 \ \Sigma_2)$$
$$P(y|x^*) \propto P(x^*|y)\,p(y)$$
$$= \mathcal{N}(\mu, \Sigma) \cdot \mathrm{Ber}(\theta)$$

training data

$x_1 \in \{\text{height}, \text{weight}\}$ $y_1 \in \{\text{gender}\}$

$\vdots$ $\qquad\qquad \vdots$

x_N $\qquad\qquad y_N$

Split data based on y positive/negative

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

$$\left\{\begin{array}{l} x_1 \quad y = 0 \\ \vdots \\ x_f \end{array}\right. \Rightarrow \text{MLE} \quad \mu_2 \ \Sigma_1$$

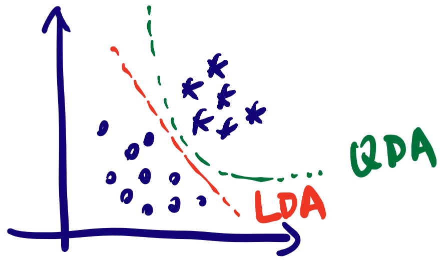
$$\theta_{MLE} = \frac{m}{m+f}$$

testing x^* $\quad p(y=1|x^*) \quad p(y=0|x^*)$

$\Sigma_1 = \Sigma_1 = \Sigma$ Linear Discriminant Analysis (LDA)

training μ_1 μ_2, θ the same QDA

Σ use all training data.



$$P(y|x) \propto P(x|y)P(y)$$

$$= \frac{1}{(2\pi)^{D/2}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \cdot P(y)$$

$$\Sigma_1 = \Sigma_2 = \Sigma$$

$$\log P(y|x) = -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$$

$$= -\frac{1}{2}\left(x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu\right)$$

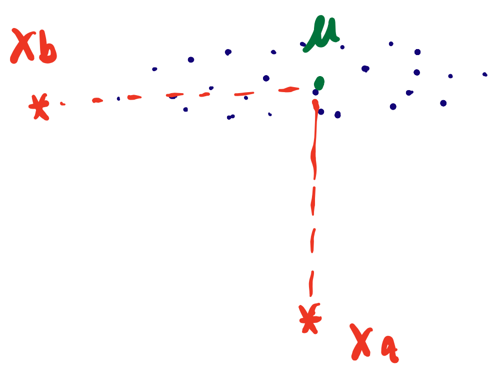
$$\log P(y=1|x^*) = -\frac{1}{2}\left(x^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1\right)$$

$$\log P(y=0|x^*) = \cdots$$

Mahalanobis distance

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

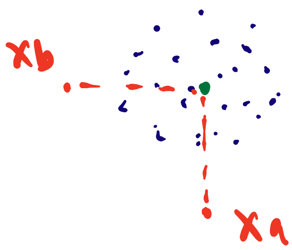
$$M(x_b) < M(x_a)$$



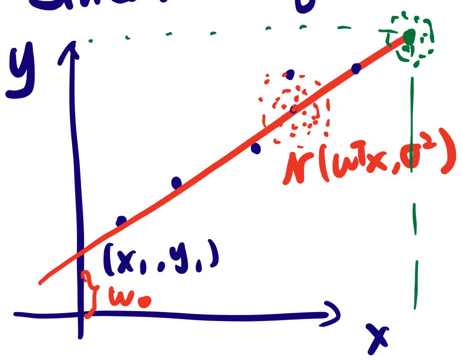
Euclidean distance

$(x-\mu)^T(x-\mu)$, special case of Mahalanobis $\Sigma = I$

$$\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$



Linear Regression



$d = 1$

$(x_1, y_1) \Rightarrow [1, x_1], y_1$

$\vdots \qquad \vdots$

$(x_n, y_n) \qquad [1, x_n], y_n$

$$y = w^T x = w_1 x_1 + w_2 x_2 + \cdots w_d x_d$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \quad w = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$$x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \quad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$ w_0 bias

$$y = w^T x = w_0 + w_1 x_1 + \cdots + w_d x_d$$

Probabilistic Interpretation

$$y = w^T x + \varepsilon$$
$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$
$$y \sim \mathcal{N}(w^T x, \sigma^2)$$

Geometric Interpretation. line fitting
least squares

$$y = w^T x$$

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

w_0: bias. intercept

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

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - w^T x_i)^2\right)$$

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

MLE, to estimate w, Set $\frac{\partial \log L}{\partial w} = 0$

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

set $y = \begin{bmatrix} y_1 \\ \vdots \\ y_i \end{bmatrix}_{N\times 1}$ $X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \end{bmatrix}_{N\times(d+1)}$ $w = \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}_{(d+1)\times 1}$

$$\log L = \boxed{(y - Xw)^T (y - Xw)}$$

$$y^T y - \underline{(Xw)^T y - y^T X w} + (Xw)^T (Xw)$$

$$\cancel{y^T y} - 2\underbrace{(y^T X w)}_{N\times 1} + w^T X^T X w$$

$$\frac{\partial \log L}{\partial w} = 0 \qquad -2(y^T X)^T + 2X^T X w = 0$$

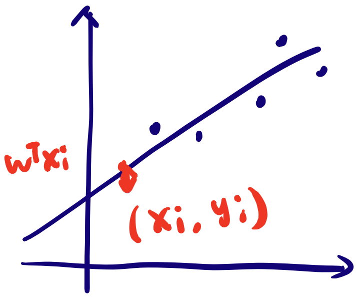
$$w = (X^T X)^{-1} X^T y$$

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

$$\sigma^2_{MLE} = \frac{1}{N}(y - Xw)^T(y - Xw)$$

Geometric Interpretation

Least Squares



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

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

$$w = (X^TX)^{-1}X^Ty$$