So far: regression

classification

→ least squares classifier

↳ binary and multi class

→ probabilistic classifiers

→ discriminative ($p(t|x)$)

(e.g logistic regression)

→ Generative ($p(x|t)$)

(e.g Gaussian Discriminant Analysis GDA/LDA)

Probabilistic classifier

→ discriminative classifiers (logistic regression)

$$p(t|x)$$

→ generative classifiers

$$p(x|t)$$

to classify a new feature vector x

$$\arg\max_t \; p(t|x) = \frac{p(x|t)\,p(t)}{\underbrace{p(x)}}$$

$$t(x) = \arg\max_t \; \overbrace{\underbrace{p(x|t)}_{x^{(i)}}}\;\underbrace{p(t)}$$

Gaussian Discriminant Analysis

$$p(x \mid t=0) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu_0)^T \Sigma^{-1} (x-\mu_0) \right)$$

$$p(x \mid t=1) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu_1)^T \Sigma^{-1} (x-\mu_1) \right)$$

$$x^{(i)} = \left(x_1^{(i)}, x_2^{(i)}, \dots, x_D^{(i)} \right) \in \mathbb{R}^D$$

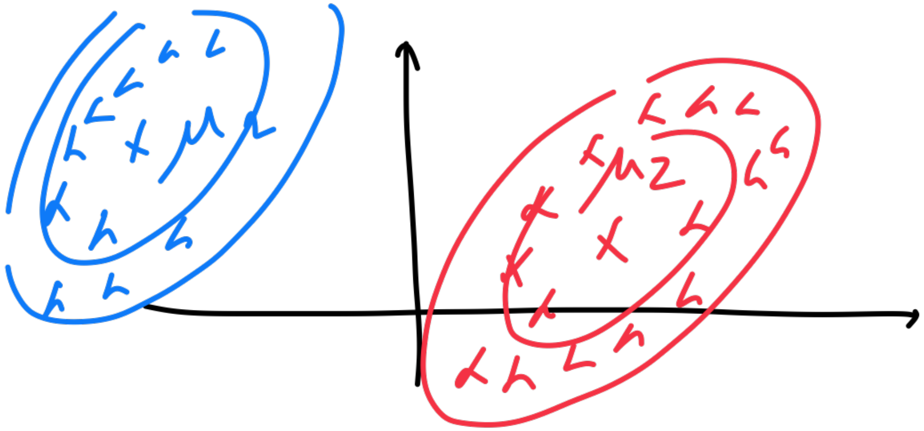
$$\left.\begin{aligned} \mu_0 &= \left(\mu_0^{(1)}, \mu_0^{(2)}, \dots, \mu_0^{(D)} \right) \in \mathbb{R}^D \\ \mu_1 &= \left(\mu_1^{(1)}, \mu_1^{(2)}, \dots, \mu_1^{(D)} \right) \in \mathbb{R}^D \end{aligned}\right\} \text{Averages of the classes}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1D} \\ \Sigma_{21} & \Sigma_{22} & & \\ \vdots & & & \\ \Sigma_{D1} & & & \Sigma_{DD} \end{bmatrix} \in \mathbb{R}^{D \times D} \quad)$$

To learn our model, we assume our training pairs $\{x^{(i)}, t^{(i)}\}$ to be independent and we take the probability of observing the set $\{x^{(i)}, t^{(i)}\}_{i=1}^{N}$ and we maximize that probability.

For that we must find the expression of $p(x, t)$ from our model of $p(x|t)$ which we can get as $p(x|t)\,p(t)$

where $p(t)$ is the class prior

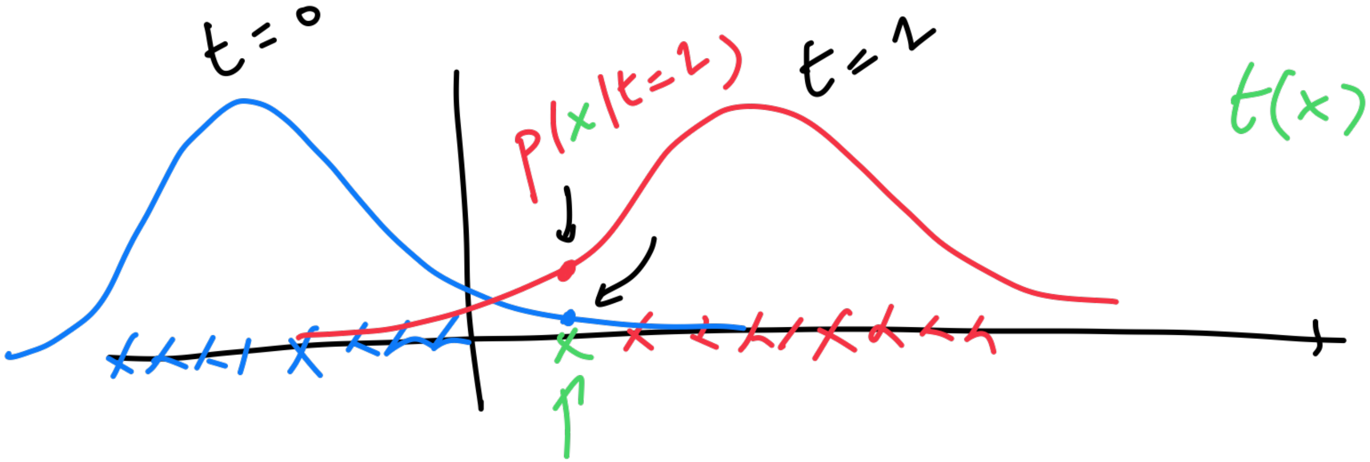


Gaussian Discriminant Analysis

binary classification

$p(x \mid t=0)$

$p(x \mid t=1)$



$$t(x) = \arg\max \{ p(x \mid t=0), \; p(x \mid t=1) \}$$

for the two classes we have

$$p(x \mid t=0) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1}(x-\mu_0)\right)$$

$$p(x \mid t=1) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)\right)$$

$$p(t) = \phi^t (1-\phi)^{1-t}$$ ϕ is unknown $\in [0,1]$

$$p(\{x^{(i)}, t^{(i)}\}) = p(x \mid t)\, p(t)$$

$$= \left(p(x \mid t=0)\, p(t=0)\right)^{1-t} \left(p(x \mid t=1)\, p(t=1)\right)^{t}$$

In this case we choose $p(t) = \begin{cases} \phi & \text{if } t=0 \rightarrow \\ 1-\phi & \text{if } t=1 \rightarrow \end{cases}$

$$= \phi^{(1-t)}(1-\phi)^{t}$$

From this, for any pair $\{x^{(i)}, t^{(i)}\}$ we get

$$p(\{x^{(i)}, t^{(i)}\}) = p(x^{(i)} \mid t^{(i)})\, p(t^{(i)})$$

$$= \left(\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_0)^T \Sigma^{-1}(x^{(i)}-\mu_0)\right) \phi \right)^{(1-t^{(i)})}$$

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

From this, we get $p(\{x^{(i)}, t^{(i)}\}_{i=1}^{N}) = \prod_{i=1}^{N} p(\{x^{(i)}, t^{(i)}\})$

and from the independence of the pairs $\{x^{(i)}, t^{(i)}\}$

L

$$
\begin{aligned}
\log \prod_{i=1}^{N} p(\{x^{(i)}, t^{(i)}\}) &= \sum_{i=1}^{N} (1-t^{(i)}) \log\left(\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\right) \\
&+ \sum_{i=1}^{N} t^{(i)} \log\left(\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\right) \\
&+ \sum_{i=1}^{N} (1-t^{(i)}) \left(-\frac{1}{2}\right) (x^{(i)}-\mu_0)^T \Sigma^{-1} (x^{(i)}-\mu_0) \rightarrow \\
&+ \sum_{i=1}^{N} t^{(i)} \left(-\frac{1}{2}\right) (x^{(i)}-\mu_1)^T \Sigma^{-1} (x^{(i)}-\mu_1) \\
&+ \sum_{i=1}^{N} (1-t^{(i)}) \log \phi + \sum_{i=1}^{N} t^{(i)} \log (1-\phi)
\end{aligned}
$$

1) Learning ϕ

taking $$\frac{\partial L}{\partial \phi} = \sum_{i=1}^{N} (1-t^{(i)}) \frac{1}{\phi} + \sum_{i=1}^{N} -t^{(i)} \frac{1}{1-\phi} = 0$$

$$\sum_{i=1}^{N} \frac{(1-\phi)(1-t^{(i)}) - t^{(i)}\phi}{\phi(1-\phi)} = 0$$

$$(1-\phi)(\overbrace{N - N_1}^{N_0}) - N_1 \phi = 0$$

where N_1 is the number of points in class C_1

$$(1-\phi) N_0 - N_1 \phi = 0$$

$$\Rightarrow \quad \phi = \frac{N_0}{N_0 + N_1} \qquad (1-\phi) = \frac{N_1}{N_0+N_1}$$

$$\frac{\partial L}{\partial \mu_0} \Rightarrow \quad \text{our 2 blue terms are of the form}$$

$$(1-t^{(i)})\left((x-\mu)^T A (x-\mu)\right) = x^T A x - \overbrace{x^T A \mu - \mu^T A x}^{2x^T A} \underbrace{+ \mu^T A \mu}_{2A\mu}$$

$$\sum_{i=1}^{N} (1-t^{(i)}) \cancel{(-2)} \cancel{\Sigma}^{-1} x^{(i)} + (1-t^{(i)}) \cancel{2} \cancel{\Sigma}^{-1} \mu_0$$

$$\sum_{i=1}^{N} (1-t^{(i)})\, x^{(i)} = \sum_{i=1}^{N} (1-t^{(i)})\, \mu_0$$

$$\sum_{i \in C_0} x^{(i)} = N_0 \mu_0$$

the solution for μ_1 follows the exact same idea

and for Σ if we use the same approach, we get

$$\Sigma = \frac{1}{N}\left(\sum_{i \in C_0} (x^{(i)}-\mu_0)(x^{(i)}-\mu_0)^T + \sum_{i \in C_1} (x^{(i)}-\mu_1)(x^{(i)}-\mu_1)^T \right)$$