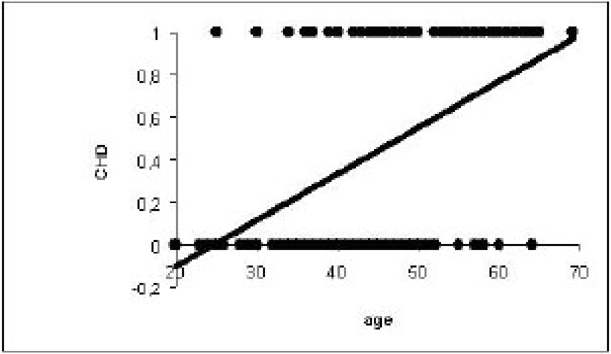
Binary classification

An example

An example

- Data : sample of 100 individuals. Two measured variables :
  - Age.
  - Suffering of a cardiac malformation (1) or not (0) Y.
- Goal of the study : has the age an influence on the fact of suffereing from this cardiac malformation.

An example



An example

Comments on the graphic

- $Y = 0$ seems to be related to youngness whereas $Y = 1$ seems to be related with oldness.
- Not sufficent to conclude
- Linear Regression : not relevant!!

Logistic Regression

The model

- Y explanatory et X explicative variable.
- Assumptions on observations : observations are independent

Logistic Regression

The model

- Assumption: conditional distribution of Y = Bernoulli distribution whose parameter $p(x)$ depends on the value x of X.
- Hence
$$\mathbb{P}(Y = y|X = x) = p(x)^y (1 - p(x))^{1-y} \ .$$

Logistic Regression

The model

- What does the function $x \mapsto p(x)$ look like?
- Properties of p :
  - it shoud take values in $[0, 1]$
  - non decreasing
- General case

$$p(x) = g({}^t\beta x)$$

$\beta \in \mathbb{R}^p$, g = linear function called the link function.

Logistic Regression

The logit model

One has

$$g(t) = \frac{\exp(t)}{1 + \exp(t)} \; .$$

Logistic Regression

Distribution log–log

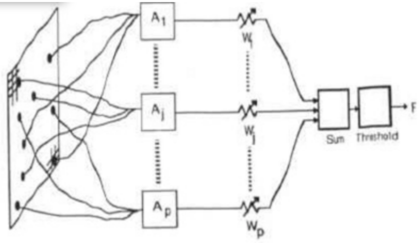
- Let
$$g(u) = 1 - e^{-e^{u}} .$$
- One has
$$p(x) = 1 - \exp(-\exp(\beta_0 + \beta_1 x)) .$$
- Caracterics of g
  - The cdf associated to g is non symmetric
  - Tails at $-\infty$ and $+\infty$ different.

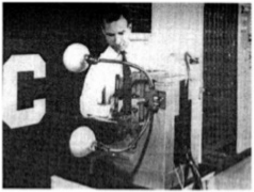
Logistic Regression

The model

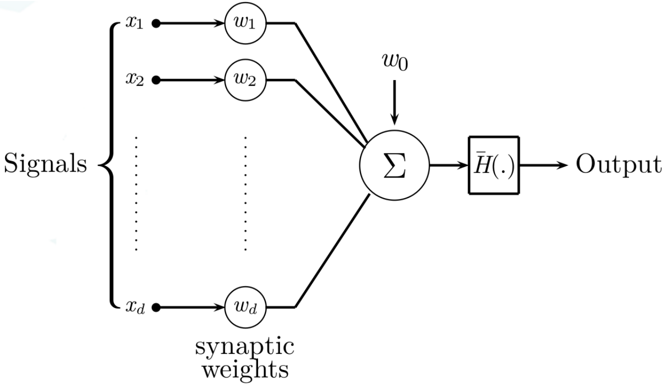
- Same shape for all these functions
- The logit model, ${}^t\beta = (\beta_0, \beta_1)$.
- Two parameters to estimate : β_0 = position, β_1 = scale.
- Properties
  - Variation : if $\beta_1 = 0$, the distribution of explanatory variable does not depend on the explicative variables.
  - if not p is strictly increasing

Le perceptron [Rosenblatt,1958]





Perceptron [Rosenblatt,1958]



Linear prediction function with parameter $w = (w_0, \overline{w}) \in \mathbb{R} \times \mathbb{R}^d$

$$h_w \quad : \quad \mathbb{R}^d \to \mathbb{R}$$
$$x \mapsto < \overline{w}, x > + w_0$$

Perceptron [Rosenblatt,1958]

Linear prediction function with parameter $w = (w_0, \overline{w}) \in \mathbb{R} \times \mathbb{R}^d$

$$
\begin{aligned}
h_w \quad : \quad & \mathbb{R}^d \to \mathbb{R} \\
& x \mapsto < \overline{w}, x > + w_0
\end{aligned}
$$

Question : can we find the parameters minimizing the distance between misclassified examples and decision boundary

schema-perceptron.png

Learning of perceptron parameters

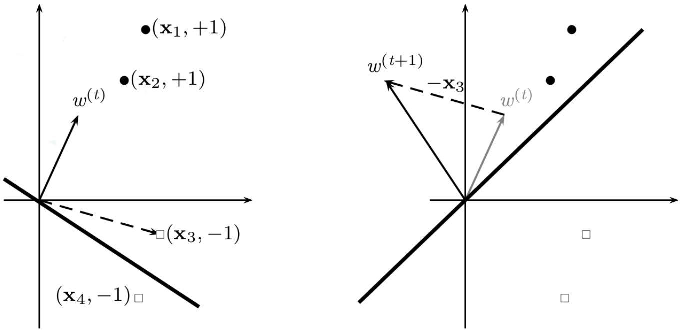
- Error measure : $L(y_1, y_2) = -y_1 \cdot y_2 1_{\{y_1 \neq y_2\}}$
- Objective function measuring empirical error

$$\widehat{L} = -\sum_{i \in \mathcal{I}} y_i \,(< \overline{w}, x_i > + w_0)$$

with $\mathcal{I}$ set of the indices of misclassified observations

- Minimization?

Learning of perceptron parameters



Perceptron algorithm

Training set $S = \{(x_i, y_i),\ i \in \{1, \cdots, m\}\}$
Initialize the weights $w(0) \leftarrow 0$
$t \leftarrow 0$
Learning rate $\eta > 0$
repeat
Choose randomly an example $(x^{(t)}, y^{(t)}) \in S$
if $y < w^{(t)}, x^{(t)} < 0$ then
$w_0 \leftarrow w_0 + \eta \times y^{(t)}$
$w^{(t+1)} \leftarrow w^{(t)} + \eta \times y^{(t)} \times x^{(t)}$
$t \leftarrow t + 1$
until $t > T$