Logistic regression

Logistic regression is used for binary classification.
It is quite similar to a simple linear regression in the sense that the objective is to find optimal weights ω to predict a variable. However, in the logistic regression we use a sigmoïd function.

Rem: "logistic" because the logistic law has a sigmoïd function as a repartition function.

Rationale behind the use of the sigmoïd function:
We look for the *à posteriori* probability $\mathbb{P}(y = 1|x) = \pi(x) = \hat{y}$.
The predicted variable $\hat{y}$ is thus a probability.

The sigmoïd function: $\sigma : z \to \frac{1}{1+e^{-z}}$ is well adapted because of two reasons:
1) We want an output variable that is included in $[0, 1]$
2) $\frac{\pi(z)}{1-\pi(z)}$ represents the relationship between a distribution and its complementary (good in binary case), and it is just a transformation of $\sigma(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{1-e^z}$

Thus, we have:
$\hat{y} = \mathbb{P}(y = 1|x) = \sigma(\omega^T x + b) = \frac{1}{1-e^{-(\omega^T x+b)}}$

Estimation
Estimation is done using maximum likelihood. Maximum likelihood is finding the parameter that maximizes the probability to have a specific event (x_i, y_i) but in our case, it is a *conditional* maximum likelihood since we want to maximize the *à posteriori* probability that depends on x.

$L(\omega, b) = \prod_{i=1}^{n} \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$

This equation has no analytic solution. We use a numeric method to find the optimal parameters (see optimizaton algorithms).