# Predicting Heart Attacks using Logistic Regression

Montaz Ali
School of Computer Science and Applied Mathematics,
University of the Witwatersrand, Johannesburg, South Africa

28th January 2021

# Outline

Mathematical Models: Linear, Polynomial, and Non-linear Regression

Linear Classification, Non-linear Transformation and Logistic Regression

Derivation of the Mathematical Optimization Model, Maximum Likelihood

The Gradient Descent Algorithm for solution

## The Generated Data Set

$$D = \{x^1, x^2, \cdots, x^N\}, x^j = (x_0, x_1, \cdots, x_n)^T, N >> n$$

Polynomial Regression: Consider the case of a single variable (say, $y$, the predictor) with the data set:

$$D = (y_1, b_1), \cdots, (y_N, b_N)$$

$$f(y) = \sum_{j=1}^{4} x_j y^{j-1} \tag{1}$$

minimizes the residual $r_i = f(y_i) - b_i$ for $i = 1 \cdots m$.

$$r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = \begin{pmatrix} f(y_1) \\ f(y_2) \\ \vdots \\ f(y_N) \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}$$

$$r = \begin{pmatrix} 1 & y_1 & y_1^2 & y_1^3 \\ 1 & y_2 & y_2^2 & y_2^3 \\ \vdots & \vdots & \vdots\vdots & \\ 1 & y_N & y_N^2 & y_N^3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}$$

$$r = Ax - b \tag{1}$$

We minimize $\|Ax - b\|^2$ which means minimizing the sum of the squared residuals $\sum_{i=1}^{N} r_i^2$, is a linear least squares problem.

$$L_m(x) = \sum_{i=0}^{n} a_i x_i = a^T x \tag{2}$$

where, $x = (x_0, x_1, \cdots, x_n)^T$, $a_0 = b$ and $x_0 = 1$.

The form like Eq(1), $\|Ax - b\|^2$, can be found and solved for $a$ and the we will have a model like $L_m(x) = a^T x$.

A nonlinear transformation, $NLT$, of $L_m(x)$ can be carried out $NLT(L_m(x)) = f(x)$, where $f(x) \in \{0, 1\}$.

Consider the sign function as your $NLT$, then

$$f(x) = sign(a^T x)$$

where $f(x) \in \{-1, 1\}$

We want to predict the probability of heart attack, hence we need $f(x) \in [0, 1]$.

Consider the Non-linear Transformation

$$F(x) = \frac{e^x}{1 + e^x}. \tag{3}$$

$$f(x) = \frac{e^{a^T x}}{1 + e^{a^T x}} \tag{4}$$

Our Noisy Target:

$$\Pr(\omega/x) = \begin{cases} f(x) & if \quad \omega = 1 \\ 1 - f(x) & if \quad \omega = -1 \end{cases} \tag{5}$$

$$\Pr(\omega/x) = F(\omega a^T x),$$

since $F(-a^T x) = 1 - F(a^T x)$

Our Data Set

$$D = \{x^1, x^2, \cdots, x^N\}$$

Consider the following likelihood function

$$L_f(a) = \prod_{i=1}^{N} \Pr\left(\omega^i / x^i\right) = \prod_{i=1}^{N} F(\omega^i a^T x^i) \tag{6}$$

$$\max_a L_f(a) = \max_a \ln \prod_{i=1}^{N} F(\omega^i a^T x^i) = \sum_{i=1}^{N} \ln\left(F(\omega^i a^T x^i)\right).$$

We know the identity

$$\max_x f(x) = -\min_x(-f(x))$$

$$\min_a \bar{L}_f(a) = \min_a -\sum_{i=1}^{N} \ln\left(F(\omega^i a^T x^i)\right)$$

$$= \min_a \sum_{i=1}^{N} \ln\left[1 + \exp\left(-\omega^i a^T x^i\right)\right] \tag{7}$$

# Optimization Procedure

$$\nabla \bar{L}_f(a) = \sum_{i=1}^{N} \frac{-\omega^i x^i}{1 + e^{\omega^i a^T x^i}}$$

The procedure is as follows:
1. Initialize $a^k$, for $k$=0, Calculate $\nabla \bar{L}_f(a^k)$,
2. Set $\alpha = 1$
3. Find $a^{k+1} = a^k - \alpha \times \nabla \bar{L}_f(a^k)$,
4. Compare if $L_f(a^{k+1}) < L_f(a^k)$ then set $k = k + 1$ and go to 5 else set $\alpha = \frac{1}{2} \times \alpha$ and go to 3.
5. Calculate $\nabla \bar{L}_f(a^{k+1})$. Stop if $\|\nabla \bar{L}_f(a^{k+1})\|$ is small else go to 2.

Thank You!