Derivadas da função de custo Erro Médio Quadrático (MSE)

1. A Função de custo:

$$J = \frac{1}{m}\sum_{i=}^{m} loss(y_i, \hat{y}_i)^2$$

2. A derivada em relação ao β:

$$\frac{\partial J}{\partial \beta} = ?$$

(1) Expandindo a função de perda:

$$J = \frac{1}{m}\sum_{i=}^{m} (\boldsymbol{y_i} - \boldsymbol{\hat{y}_i})^2$$

(2) Expandindo o cálculo da predição:

$$J = \frac{1}{m}\sum_{i=}^{m} (y_i - (\boldsymbol{\vec{x_i}\beta} + \boldsymbol{\alpha}))^2$$

(3) Distribuindo o expoente:

$$J = \frac{1}{m}\sum_{i=}^{m} [\boldsymbol{y_i}^{\boldsymbol{2}} - \boldsymbol{2y_i}\,(\boldsymbol{\vec{x_i}\beta} + \boldsymbol{\alpha}) + (-\boldsymbol{\vec{x_i}\beta} - \boldsymbol{\alpha})^{\boldsymbol{2}}]$$

(4) Distributiva no segundo termo e no expoente do terceiro:

$$J = \frac{1}{m}\sum_{i=}^{m} [y_i{}^2 - \boldsymbol{2y_i\,\vec{x_i}\beta} - \boldsymbol{2y_i\,\alpha} + (\,(-\boldsymbol{\vec{x_i}\beta})^{\boldsymbol{2}} + \boldsymbol{2\vec{x_i}\beta\alpha} + (-\boldsymbol{\alpha}^{\boldsymbol{2}})]$$

(5) Calculando derivadas parciais para cada elemento. Lembrem que elementos onde não aparece $\boldsymbol{\beta}$ são considerados constantes, logo

suas derivadas são 0, onde há $\boldsymbol{\beta}$ a derivada torna-se 1, e onde há $\boldsymbol{\beta}^2$ a derivada torna-se $2\boldsymbol{\beta}$:

$$\frac{\partial J}{\partial \beta} = \frac{1}{m}\sum_{i=}^{m}[\cancel{y_i{}^2} - 2\vec{x_i}y_i \cancel{- 2y_i\alpha} + 2\vec{x_i}^2\beta + 2\vec{x_i}\alpha \cancel{- \alpha^2}]$$

(6) Reorganizando os termos por fator comum (colocando $-2\vec{\boldsymbol{x_i}}$ em evidência):

$$\frac{\partial J}{\partial \beta} = \frac{1}{m}\sum_{i=}^{m}[-\boldsymbol{2\vec{x_i}(y_i - \overrightarrow{x_i\beta} - \alpha)}]$$

$$= \frac{1}{m}\sum_{i=}^{m}[-2\vec{x_i}(y_i - (\boldsymbol{\vec{x_i}\beta + \alpha}))]$$

3. A derivada em relação ao α:

$$\frac{\partial J}{\partial \alpha} = ?$$

(1) Partindo da função J já com a distributiva aplicada:

$$J = \frac{1}{m}\sum_{i=}^{m}[y_i{}^2 - 2y_i\vec{x_i}\beta - 2y_i\alpha + ((-\vec{x_i}\beta)^2 + 2\vec{x_i}\beta\alpha + (-\alpha^2)]$$

(2) Calculando derivadas parciais para cada elemento. Lembrem que elementos onde não aparece α são considerados constantes, logo suas derivadas são 0, onde há α a derivada torna-se 1, e onde há α^2 a derivada torna-se 2α:

$$\frac{\partial J}{\partial \alpha} = \frac{1}{m}\sum_{i=}^{m}[\cancel{y_i{}^2} \cancel{- 2y_i\vec{x_i}\beta} - 2y_i + (\cancel{(-\vec{x_i}\beta)^2} + 2\vec{x_i}\beta + 2\alpha^1)]$$

(3) Reorganizando os termos por fator comum (colocando -2 em evidência):

$$\frac{\partial J}{\partial \alpha} = \frac{1}{m}\sum_{i=}^{m}[-2(y_i - \vec{x_i}\beta - \alpha)]$$

$$= \frac{1}{m}\sum_{i=}^{m}[-2(y_i - (\boldsymbol{\vec{x_i}\beta + \alpha}))]$$