1 Calculate vector calculus $\frac{\partial \mathcal{L}}{\partial \mathcal{W}} = (X^T(\hat{y} - y))$

Suppose

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} \\ 1 & \dots & \dots \\ 1 & x_1^{(n)} & x_2^{(n)} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$\hat{y} = \sigma(Xw)$$

L = -$(y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i))$

Apply the Chain rule:

$$\frac{\partial L}{dW} = \frac{\partial L}{\partial \hat{y}_i} * \frac{\partial \hat{y}_i}{\partial W} \frac{\partial L}{\partial \hat{y}_i} = -X \frac{\partial (y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i))}{\partial \hat{y}_i} =$$

$$-X \left(\frac{y_i}{\hat{y}_i} - \frac{1 - y_i}{(1 - \hat{y}_i)} \right) * \hat{y}_i * (1 - \hat{y}_i) = -X((y_i * (1 - \hat{y}_i) - (1 - y_i) * \hat{y}_i)) = X(\hat{y}_i - y_i)$$

$$\longrightarrow \frac{\partial L}{\partial w} = X^T * (\hat{y} - y), X^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(n)} \end{bmatrix}$$

2 Use Gradient Descent, implement Logistic Regression

Python FIle

3 Run Logistic Regression

Python FIle

4 Draw boundary line

Python FIle

5 Prove that with Logistic model, Loss Binary Cross Entropy is convex function of W whereas Loss Mean Square Error is not

5.1 Loss Binary Cross Entropy

$$J(w) = -\frac{1}{m} \left(y^T \log \sigma(Xw) + (1 - y)^T \log \sigma(-Xw) \right)$$

where each row of X is one of our training examples and we made use of some identities introduced along with the logistic function. Using some elements of matrix calculus, one can show that the gradient of our loss function with respect to w is given by

$$\nabla_w J(w) = \frac{1}{m} X^T(\sigma(Xw) - y)$$

Similarly, the Hessian matrix reads

$$H(w) = X^T D X$$

with

$$D = \frac{1}{m} \operatorname{diag}(\sigma(Xw)(1 - \sigma(Xw))$$

From this point, one can easily show that

$$\begin{aligned} a^T H(w) a &= a^T X^T D X a \\ &= a^T X^T D^{1/2} D^{1/2} X a \\ &= \left\| D^{1/2} X a \right\|_2^2 \\ &\geq 0 \\ &\longrightarrow Convex \end{aligned}$$

5.2 Loss MSE

- Using logistic regression

$$\sigma = \frac{1}{1 + \exp\left(-\beta_0 - \beta_1 \cdot x\right)}$$

for the MSE cost function

$$h = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sigma_{\beta_0, \beta_1}\left(x_i\right)\right)^2$$

gives

$$h = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sigma_i\right)^2$$

- Calculate first order derivative of summand

$$\frac{\partial \left(y_i - \sigma_i\right)^2}{\partial \beta_k} = -2\left(y_i - \sigma_i\right) \frac{\partial \sigma_i}{\partial \beta_k}$$

- The last factor is according to chain rule

$$\frac{\partial \sigma}{\partial \beta_k} = \frac{\partial \sigma}{\partial \operatorname{lm}} \frac{\partial \operatorname{lm}}{\partial \beta_k}$$

- Calculate second order derivative is

$$\frac{\partial^2 \left(y_i - \sigma_i\right)^2}{\partial \beta_k^2} = -2\left(y_i - 2 y_i \sigma_i - 2\sigma_i + 3\sigma_i^2\right) \cdot \sigma_i \cdot \left(1 - \sigma_i\right) \left(\frac{\partial \operatorname{lm}}{\partial \beta_k}\right)^2$$

a) $y_i = 0$

$$\begin{aligned} &- \quad 2\left(-2\sigma_i + 3\sigma_i^2\right) \quad \cdot \quad g \\ &= -2\left(-2 + 3\sigma_i\right)\sigma_i \cdot g \end{aligned}$$

change of sign $\longrightarrow \sigma_i \in [0, 1]$

b) $y_i = 1$

$similarly$

To conclude, $\frac{\partial^2 h}{\partial \beta^2} \geq 0$ can not happen
$\longrightarrow$ Not convex