Note: This is the summary note from Udacity Introduction to Deep Learning with PyTorch

Logistic Regression

- The differences between linear-regression and logistic-regression.

Comparison Chart

BASIS FOR COMPARISON	LINEAR REGRESSION	LOGISTIC REGRESSION
Basic	The data is modelled using a straight line.	The probability of some obtained event is represented as a linear function of a combination of predictor variables.
Linear relationship between dependent and independent variables	Is required	Not required
The independent variable	Could be correlated with each other. (Specially in multiple linear regression)	Should not be correlated with each other (no multicollinearity exist).

https://techdifferences.com/difference-between-linear-and-logistic-regression.html

- Algorithm:
  - Take your data
  - Pick a random model
  - Calculate error
    - Binary classification problem

$$Cross\ Entropy = -\sum_{i=1}^{m} y_i ln(p_i) + (1 - y_i)\, ln(1 - p_i)$$

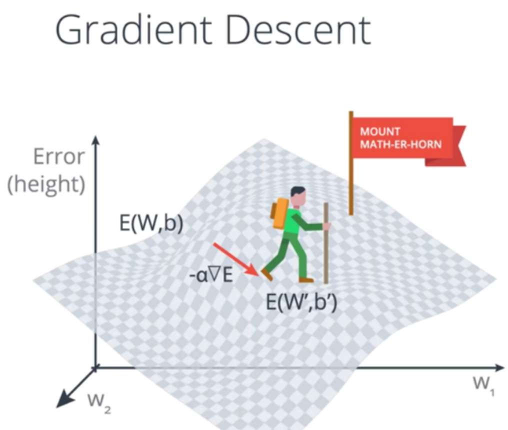
    - Multi classification problem

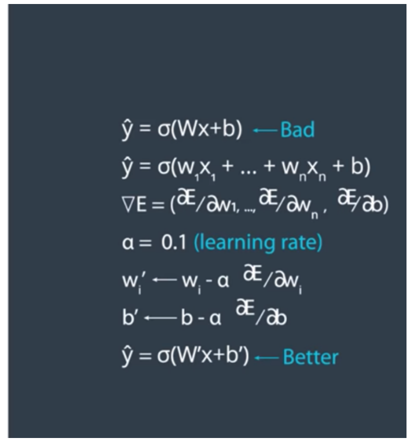
$$Cross\ Entropy = -\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} ln\left(p_{ij}\right)$$

  - Minimize error; E = error function
    - E(W, b) → Use gradient decent → we get new E(W', b') smaller error function

Gradient Descent

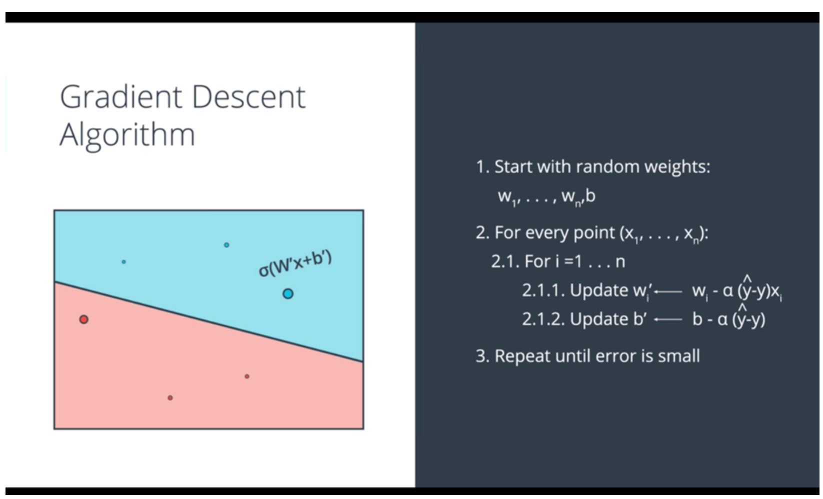
- Math:





- Derivative of sigmoid function : $\sigma'(x) = \sigma(x)(1 - \sigma(x))$
- This function come from derivation of sigmoid function. Detail in https://math.stackexchange.com/questions/78575/derivative-of-sigmoid-function-sigma-x-frac11e-x

Gradient Descent Algorithm



- Math : article : https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e
- Repeat until fixed number = epoch
- Similar to perceptron algorithm ???

Mini Summary

- Important functions

  - Sigmoid activation function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

  - Output (prediction) formula

$$\hat{y} = \sigma(w_1 x_1 + w_2 x_2 + b)$$

  - Error function

$$\boldsymbol{Error}(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$$

  - The function that updates the weights

$$w_i \longrightarrow w_i + \alpha(y - \hat{y}) x_i$$

$$b \longrightarrow b + \alpha(y - \hat{y})$$

- Problem: logistic regression vs linear regression
- Logistic regression function
  - Calculate error
    - Binary classification problem

$$Cross\ Entropy = -\sum_{i=1}^{m} y_i ln(p_i) + (1 - y_i)\, ln(1 - p_i)$$

    - Multi classification problem

$$Cross\ Entropy = -\sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} ln\left(p_{ij}\right)$$

  - Minimize error; E = error function
    - $\mathrm{E(W, b)}$ → Use gradient decent → we get new $\mathrm{E}(W', b')$ smaller error function