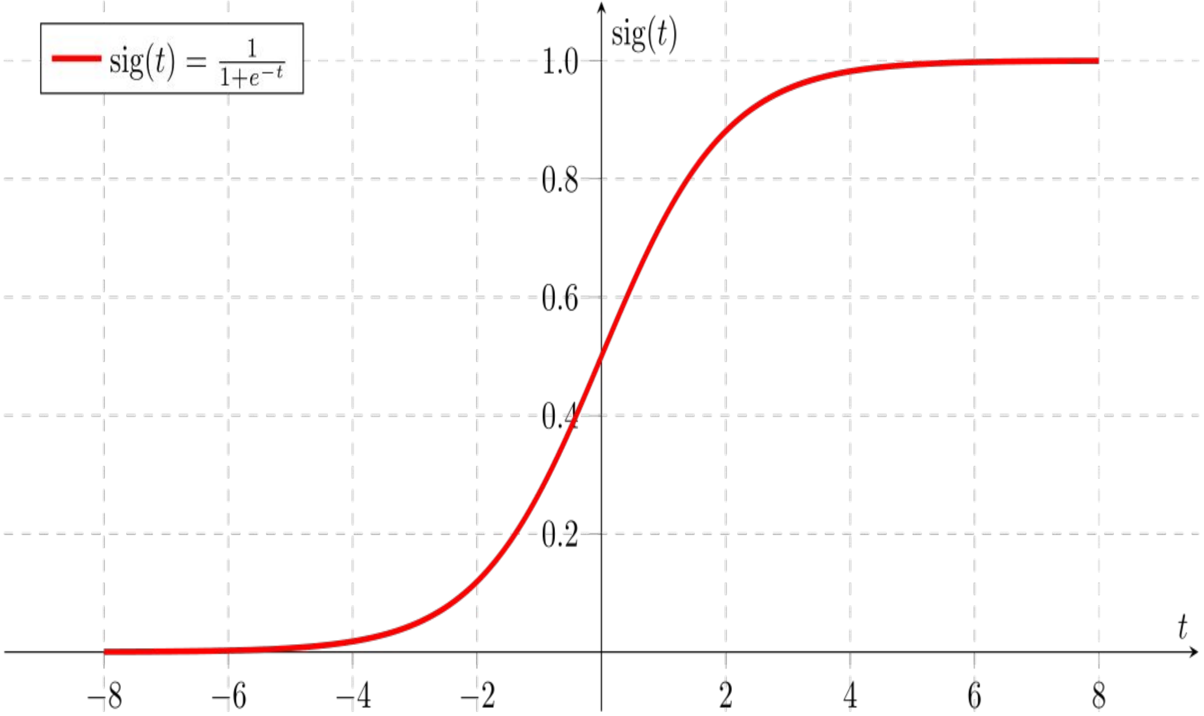
Logical Rhythm

Class #2: 27th August 2019

Logistic Regression (Classification Algorithm)



$\mathrm{sig}(z) = 1/(1+e^{-z})$

$Y = \mathrm{sig}(\sum w_i x_i)$

"*Probability of output to be categorically 1 for given value of x.*"

<u>Reasons for not using Linear Regression</u> -

Value not in finite range.

Works with luck for best fit.

Logistic regression cost function

0

If actual y=1

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$= -\frac{1}{m}\Big[\sum_{i=1}^{m} \boxed{y^{(i)} \log h_\theta(x^{(i)})} + \boxed{(1-y^{(i)}) \log\left(1-h_\theta(x^{(i)})\right)}\Big]$$

0 If actual y=0

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Note: $y = 0$ or 1 always

Bayes’ Theorem

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

Derivation

- P(A and B) = P(A)*P(B|A)
- P(A and B) = P(B)*P(A|B)
- P(A)*P(B|A) = P(B)*P(A|B)
- **P(B|A) = P(B)*P(A|B)/P(A)**

Multivariate Naive Bayes

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

$$X = (x_1, x_2, x_3, \ldots.., x_n)$$

$$P(y|x_1, \ldots, x_n) = \frac{P(x_1|y)P(x_2|y)\ldots P(x_n|y)P(y)}{P(x_1)P(x_2)\ldots P(x_n)}$$

$$P(y|x_1, \ldots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y)$$

Additional Information

- **Pros and Cons**
  - **Pros**
    - Fast and accurate when independent features are involved
    - Works better for categorical valued features, as numeric features involved normal distribution assumption
  - **Cons**
    - Features are rarely independent in real-life problems
    - **Zero frequency** problem

- **Applications:** Recommender System, Text Classification, Sentiment Analysis