

Linear Regression & Logistic Regression

By NtD

Review

Vector is way of representation of data in numerical format.

["Hello", "world", "!"] \Rightarrow [103, 98, 22]

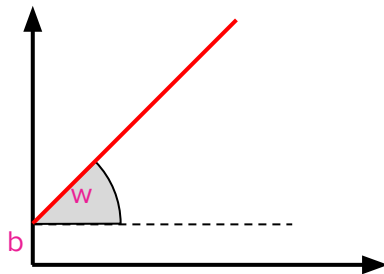
["Male", 232.0, 12000, "Yes"] \Rightarrow [1, 232.0, 12000, 1]

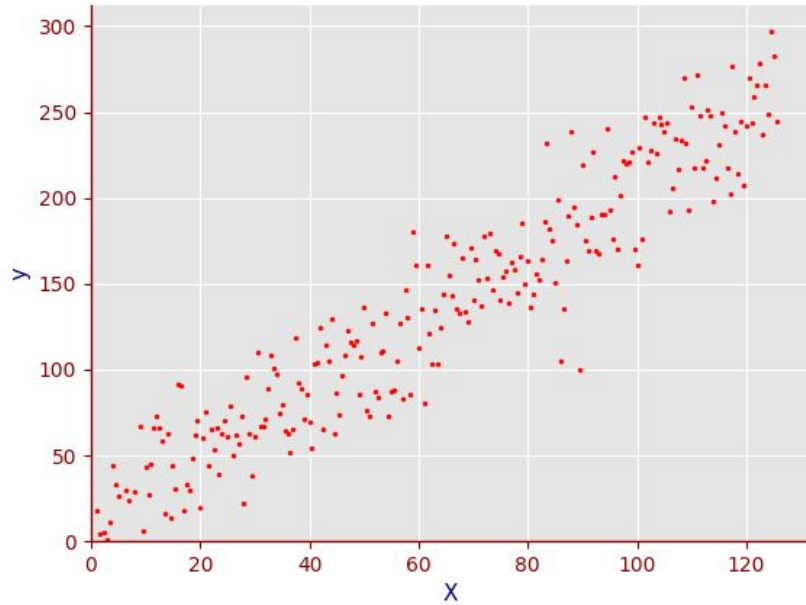
Linear Regression

Linear function: $f(x) = \mathbf{w}x + \mathbf{b}$

w: weight, coefficient

b: bias, intercept



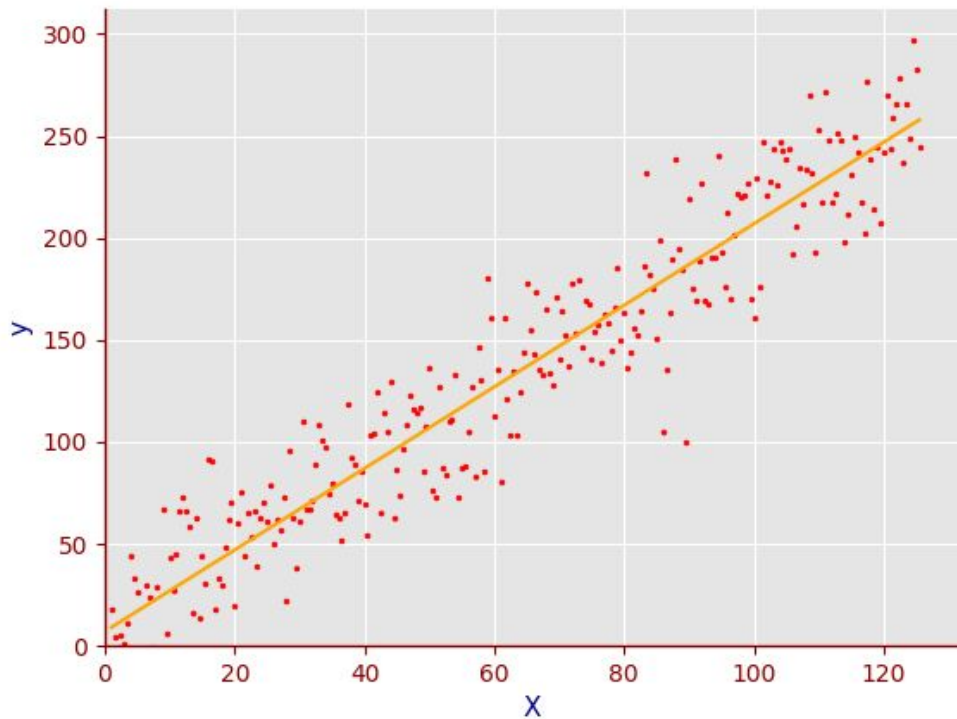


Which function ($w?$, $b?$)
this data is representing
for?

$$w = 2$$

$$b = 7$$

$$f(x) = 2x + 7$$



Metrics

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

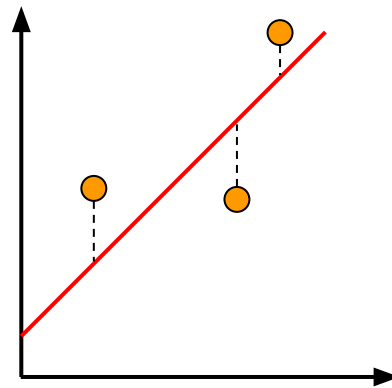
$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

$$RMSE = \sqrt{MSE}$$

MAE: Mean Absolute Error

MSE: Mean Squared Error

RMSE: Root Mean Squared Error



Usability and Drawbacks

Usability:

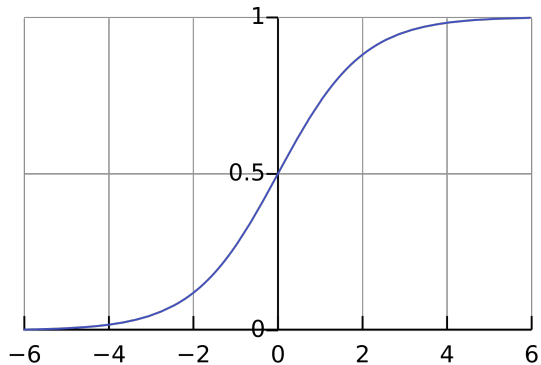
- > Used widely in economics and finance
- > Predict or forecast trends in short or mid-term
- > Linear algebra is the foundation for mathematics as well as machine learning/deep-learning

Drawbacks:

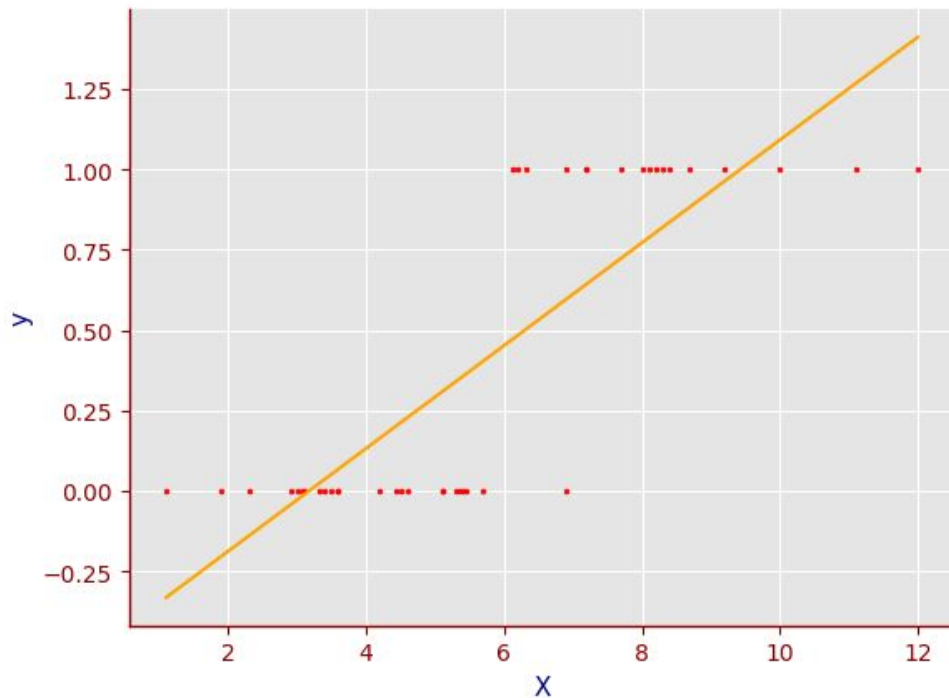
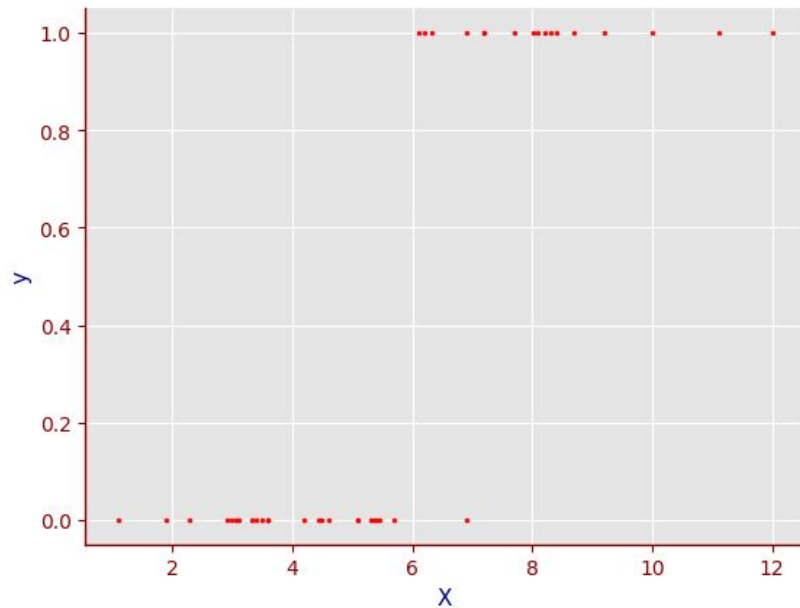
- > Unfit non-linear or complex datasets

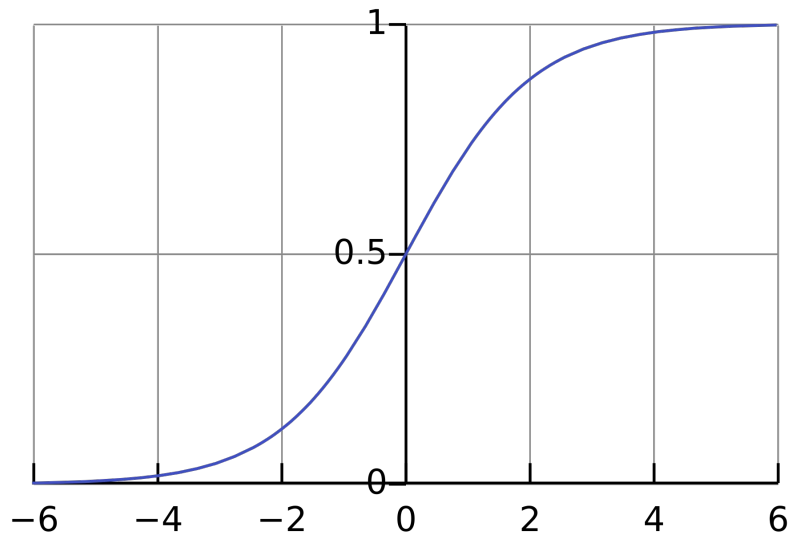
Logistic Regression

For classification problems



Problem of Linear Regression



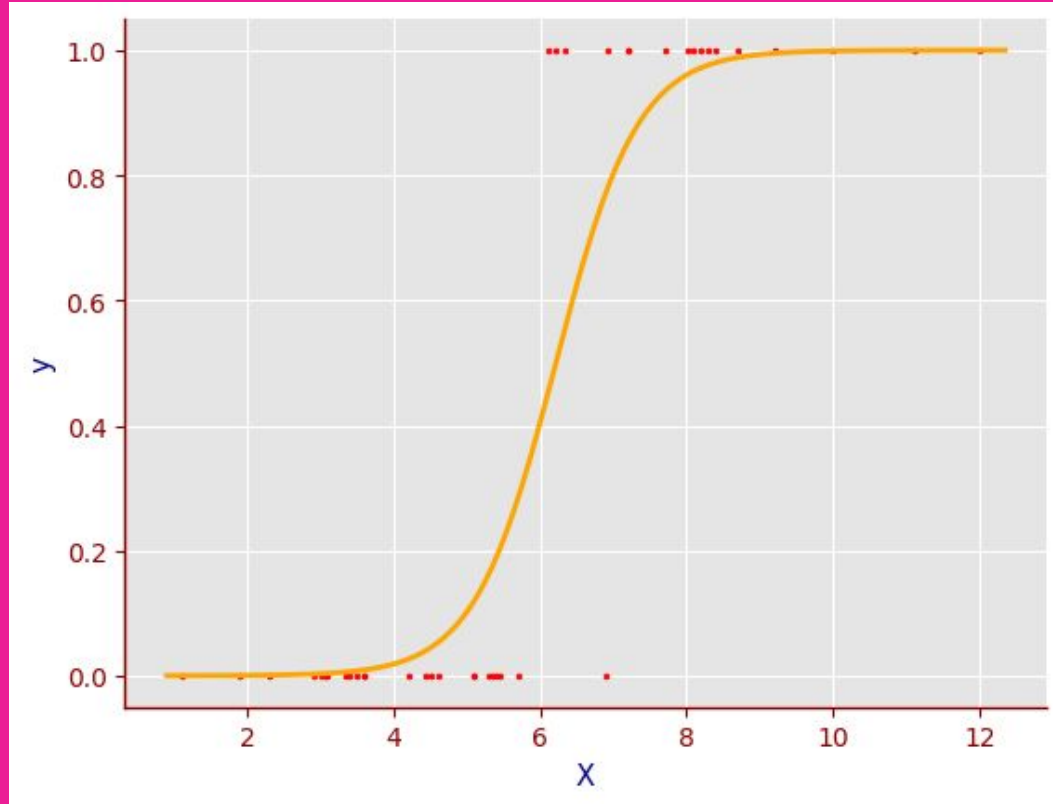


Logistic or Sigmoid Function

https://en.wikipedia.org/wiki/Sigmoid_function

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

Applied Logistic Regression



Metrics: Confusion Matrix

True Negative (TN)	False Positive (FP)
False Negative (FN)	True Positive (TP)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \left(\frac{Precision * Recall}{Precision + Recall} \right)$$

Usability and Drawbacks

Usability:

- > Effective for linear-classification: customer churn, frauds detection, quality control...
- > Simple, fast, yet effective
- > Logistic curve can be used in many use cases and in deep-learning

Drawbacks:

- > Do not work well with imbalanced classification
- > Unfit complex datasets

Libraries and Frameworks

Library	Description
pandas	Data visualization and preprocessing
numpy	Vectorization and processing data
matplotlib, seaborn	Plotting and data visualization
scikit-learn	Data processing and machine learning models
xgboost	machine learning models
jupyter	Jupyter notebook server

Thank you!