Linear Regression

- Linear Regression is one of the supervised algorithms that is used to predict the continuous values based on the given inputs.
- It follows the mathematical equation given as:

  y=w1x1+w2x2+⋯+wn.xn+b

  x is the given input features.

  y is the value to be predicted.

  w describes the importance of each feature.

  b represents the bias. (intercept)

  n is the number of input features.
- Graphically it represents a straight line.

WHY? Linear Regression

- To determine the best values of weight and biases such that the difference between the actual value and the predicted value remains low.
- To measure how good, the model performs, we use loss function, Mean Square Error(MSE) used for Linear Regression:

  MSE=1/m[$\sum_{i=1}^{m}(yi - yp)$^2]

Where , yi=actual values

yp=predicted values

- It follows linear relationship between the target and the features.

Limitations

- Only suitable for continuous datas.
- Fails to work for the categorical datas or classification purpose.
- Sensitive to external noise,may effect output .

Logistic Regression

- It is other supervised learning ML algorithm, that overcomes the above mentioned limitations of Linear Regression.
- It is used to calculate the probability between two classes, either 0 or 1.
- It follows the mathematical equation:

  y=w1x1+w2x2+⋯+wn.xn+b

  - It then applies sigmoid function so the equation becomes:
  - **σ(z)=1/(1+e^(−z))**
  - **P(y=1)y=1/(1+e^(−(w1x1+w2x2+⋯+b))**

  If **y>0.5** the predicted class is **1.**

  If **y<0.5** the predicted class is **0.**

Sigmoid function is also known as activation function for binary class where as **Softmax** Logistic Regression

- It is other supervised learning ML algorithm, that overcomes the above mentioned limitations of Linear Regression.
- It is used to calculate the probability between two classes, either 0 or 1.
- It follows the mathematical equation:

  y=w1x1+w2x2+⋯+wn.xn+b

  - It then applies sigmoid function so the equation becomes:
  - **σ(z)=1/(1+e^(−z))**
  - **P(y=1)y=1/(1+e^(−(w1x1+w2x2+⋯+b))**

  If **y>0.5** the predicted class is **1.**

  If **y<0.5** the predicted class is **0.**

  - **Sigmoid function** is also known as **activation function.** It is used for binary class whereas for multiple classes, the activation function used is **Softmax Function.**
- It first predicts the log-odds before predicting the probability.
- Log odds refers to the ratio of an event occurring to that of an event not occurring.
- Mathematically:

log-odds=log(p/1−p)

- Logistic Regression uses Binary Cross Entropy for the loss function, whereas for multiple classes it uses sparse categorical entropy.
- It first predicts the log-odds before predicting the probability.
- Log odds refers to the ratio of an event occurring to that of an event not occurring.
- Mathematically:

  log-odds=log(p/1−p)
- Logistic Regression uses Binary Cross Entropy for the loss function, whereas for multiple classes it uses **Categorical Cross Entropy.**