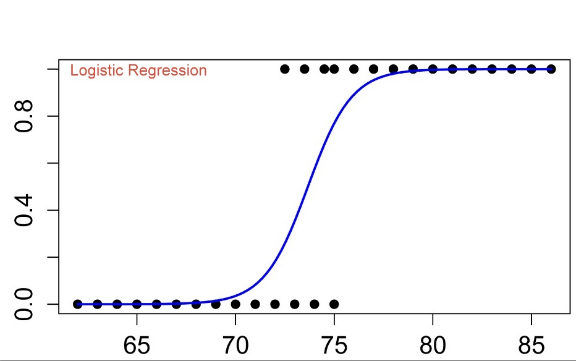
**Machine Learning:**

1. **Supervised** [*Machine learns under guidance. E.g. Teacher guiding the students. Machine learns by feeding the labelled (weight, height, age, color etc.) data. Input and process is given, output is to be computed.*]
   1. Classification [*predicting label/class*]
      1. Binary [*E.g. classifying the emails into two categories viz. spam and non-spam*]
      2. Multiclass [*E.g. classifying the non-spam emails into categories viz. primary, Promotions, Social and Updates*]
   2. Regression [*predicting continuous quantity/variable. E.g. weight of a person. E.g. Given Gold rate data, predict gold price (dependent variable) at a given time (independent variable).*]
      1. Linear [*Predicting dependent variable [continuous] based on independent variable. Dependent variable is always continuous.*]
      2. Logistic [*To solve classification problem. To predict dependent variable [categorical, e.g. 1/0, Yes/No, True/False] from given independent variables*]
2. **Unsupervised** [*No labelled data, no guide. Machine has to figure out label among data, and pattern in data, on its own. Machine makes prediction about the output.*]
3. **Reinforcement** [*Machine does not have any idea about the data, label, pattern etc. It tries to learn on trial-and-error/hit-and-trial basis. Learn from the experience.*]

**Logistic Regression:**

Sigmoid / S Curve: We need to set a threshold (E.g. 0.5). And any value between 0 to threshold belongs to or classified as ‘0’ and any value between threshold to 1 belongs to or classified as ‘1’. Y value (dependent variable) can be any continuous value between 0 to 1 whereas X value (independent variable) can be any categorical value, either 0 or 1.



= Constant C + B1X1 + B2X2 + B3X3 + …

Y : Probability of an event to happen [Dependent categories. *e.g. 1/0, Yes/No, True/False*]

X1, X2, X3 : [Continuous] Independent variables

B1, B2,B3 : Respective coefficient of X1, X2, X3

**Linear Regression Use Case:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Marketing Expense** | **Sales** | | 500 | 200 | | 600 | 300 | | 1200 | 500 | | 1500 | 700 | | 2000 | 600 | | 1000 | 650 | | 500 | 350 | |  |

**Observations**: Positive Correlation. That is, with the increase in marketing expense, the sales increase.

**Logistic Regression Use Case:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Previous Round Result** | **Qualify for next round** | | 8.5 | 1 | | 3.6 | 0 | | 4.9 | 0 | | 7.7 | 1 | | 6.7 | 1 | | 5.8 | 1 | | 4.4 | 0 | | 9.1 | 1 | | **Threshold** | **5.5** | |  |

Sigmoid or Logit Function or S Curve z =

e = Euler’s number = 2.71828

z = mx + c [Linear regression, Straight Line]

Application: To detect fraudulent email/traffic. Image classification.

**Comparison:**

|  |  |  |
| --- | --- | --- |
|  | **Linear Regression** | **Logistic Regression** |
| Definition |  |  |
| Variable Type | Continuous dependent | Categorical dependent |
| Estimation | Least square | Maximum likelihood |
| Equation |  |  |
| Best fit line | Straight line | S Curve |
| Outcome | Integer value | Binary value |
| Applications |  |  |