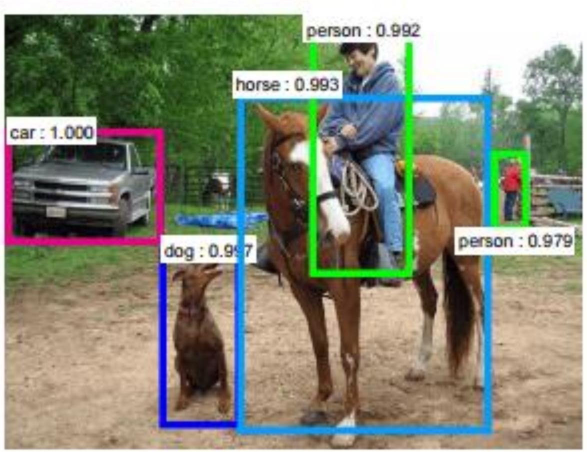
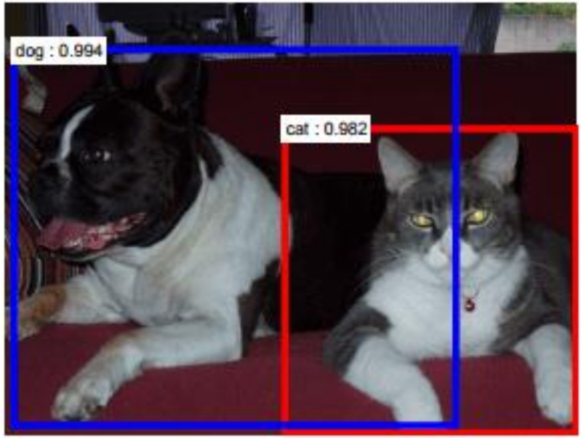
Classification

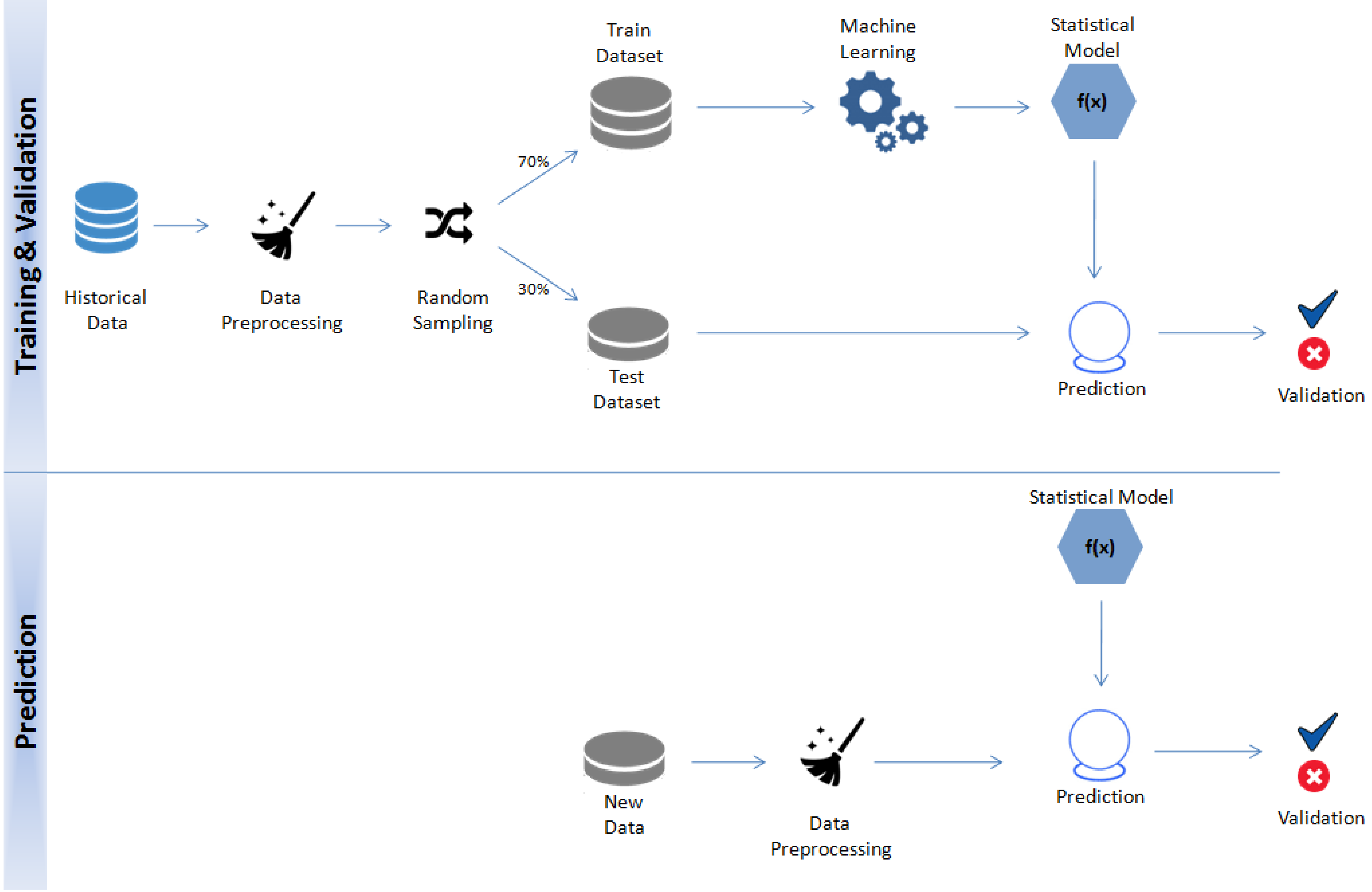


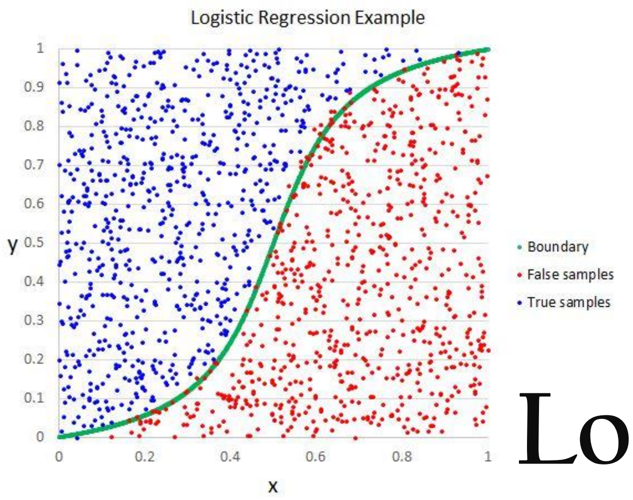


INNOMATICS TECHNOLOGY HUB

Domain	Question
Telecom	Is a customer likely to leave the network? (churn prediction)
Retail	Is he a prospective customer? that is likelihood of purchase vs. non-purchase?
Insurance	To issue insurance should a customer be sent for a medical checkup?
Insurance	Will the customer renew the insurance?
Banking	Will a customer default on the loan amount?
Banking	Should a customer be given a loan?
Manufacturing	Will the equipment fail?
Health Care	Is the patient infected with a disease?
Health Care	What type of disease does a patient have?
Entertainment	What is the genre of music?

INNOMATICS TECHNOLOGY HUB


Training & Validation
Historical Data
Data Preprocessing
Random Sampling
70%
30%
Train Dataset
Test Dataset
Machine Learning
Statistical Model
f(x)
Prediction
Validation
Prediction
Statistical Model
f(x)
New Data
Data Preprocessing
Prediction
Validation



Logistic Regressions

(Classification)

Simple Linear Regression overview

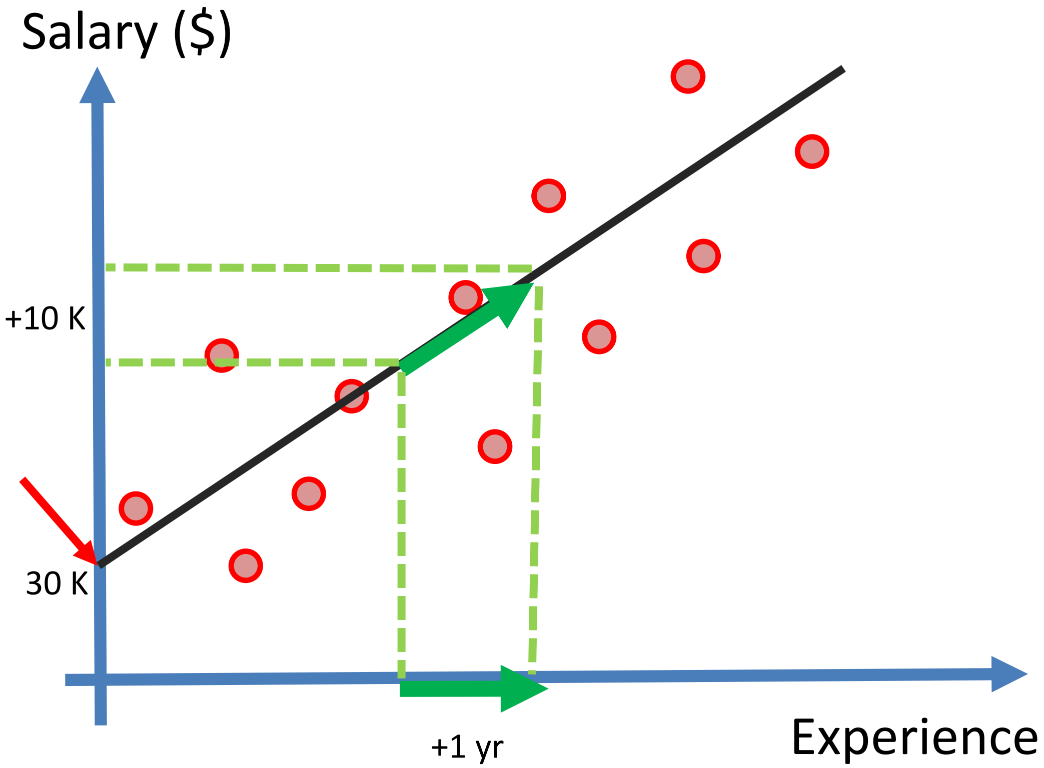
Simple Linear Regression

$$y = b_0 + b_1 * x_1$$

Multiple Linear Regression

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_n * x_n$$

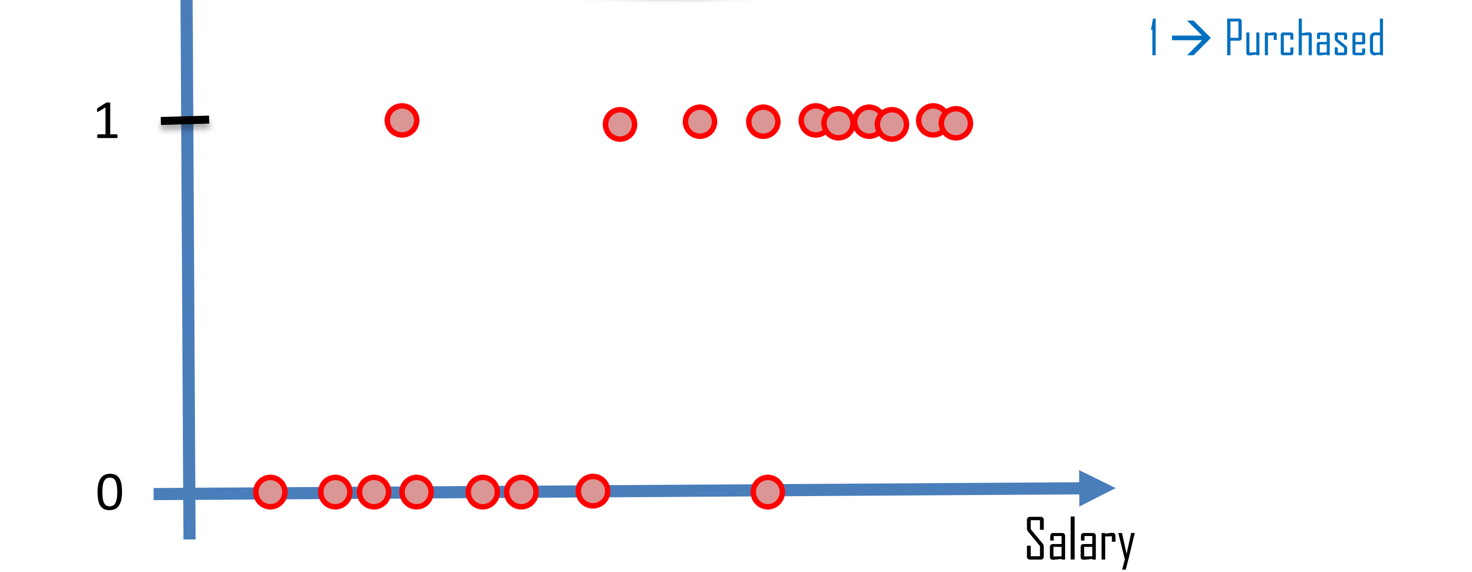
Simple Linear Regression:



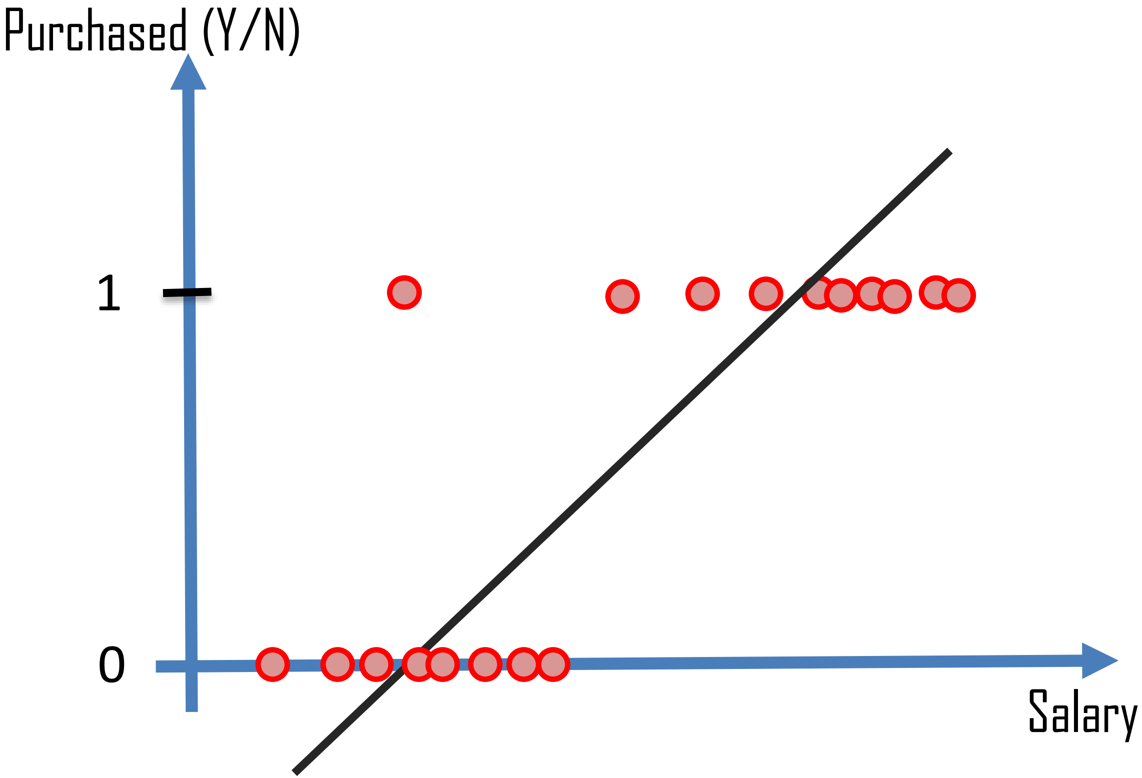
$$y = b_0 + b_1 * x_1$$

$$Salary = b_0 + b_1 * Experience$$

Retail (Like hood t urchase)

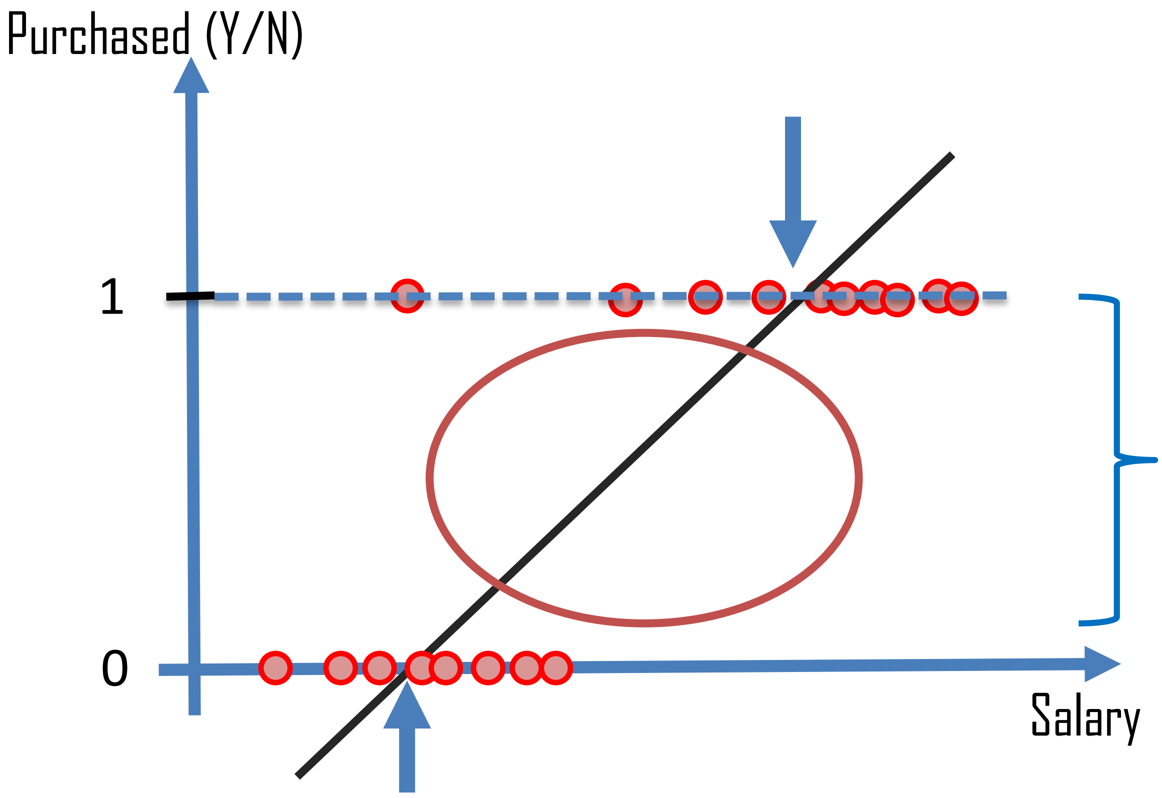


Linear Regression Prediction in Retail (Like hood to purchase)



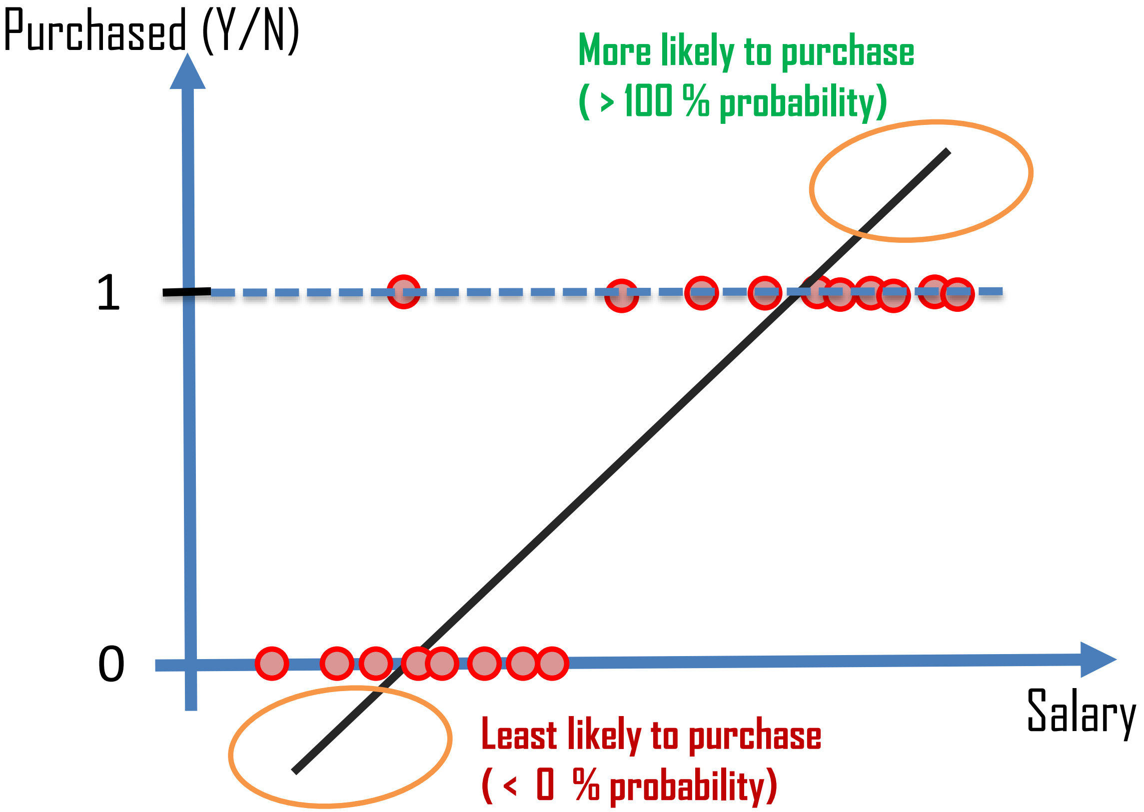
0 → Non-Purchased
1 → Purchased

Probability of Likely hood to purchase



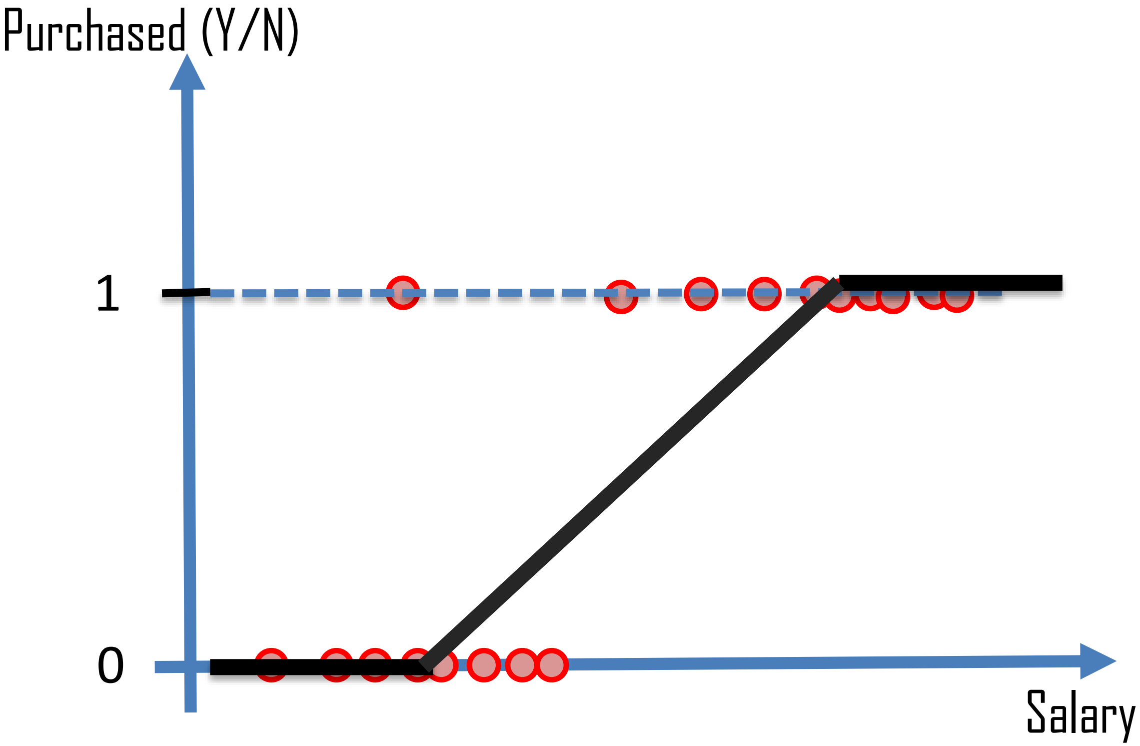
Gives probability scores,
- As "Salary" increases there is more likely hood that he/she can purchase a CAR

Probability of Likely hood to purchase



The Line above and below the probability limit is telling that there is more likely to purchased and not-purchased

Logistic Regression



Logistic Regression (Scientific Approach)

$y = b_0 + b_1 \; * x_1$

$$p = \frac{1}{1 + e^{-y}}$$

$\log(\frac{p}{1-p}) = \boldsymbol{b_0 + b_1 \; * x_1}$

INNOMATICS TECHNOLOGY HUB

Logistic Regression (Scientific Approach)

- Fundamental Idea is to introduce **sigmoid** or **Logit function** to regression equation.

Eq. of linear regression: $y = b_0 + b_1 * x_1$

Logistic regression can be explained better in odd ratios.

"Odd of an event occurring are defined $= \dfrac{\text{probability of event occurring}}{\text{probability of event } \textbf{not} \text{ occurring}}$"

Logits = log(Odd ratio)

Logistic Regression (Scientific Approach)

$$odd\ ratio\ of\ purchased\ vs\ not\ purchased = \frac{P(y=1)}{1\ - P(y-1)}$$

$$logits = y = \log_e\left(\frac{P(y=1)}{1\ - P(y-1)}\right)$$

From Linear Regression

$$y = b_0 + b_1\ * x_1$$

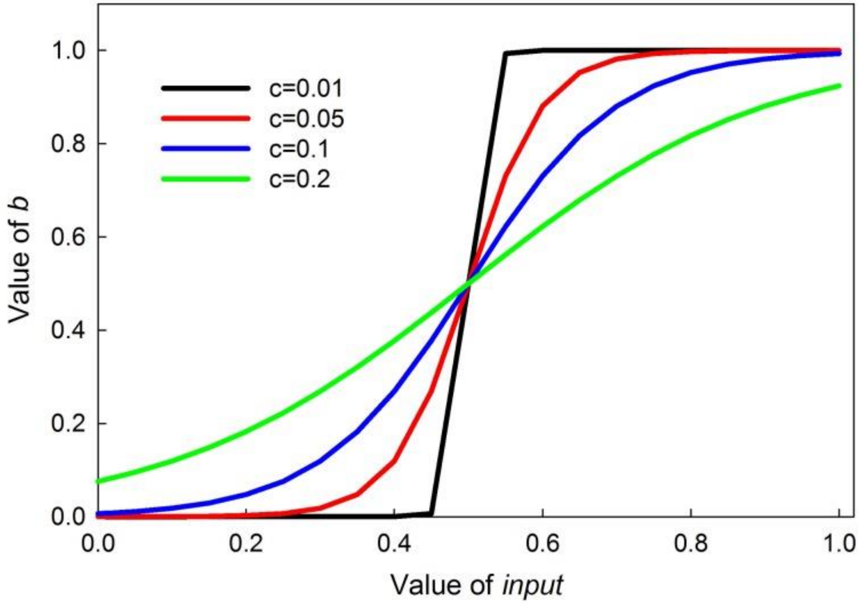
$$\log_e\left(\frac{P}{1-P}\right) = b_0 + b_1\ * x_1$$

INNOMATICS TECHNOLOGY HUB

Logistic Regression (Scientific Approach)

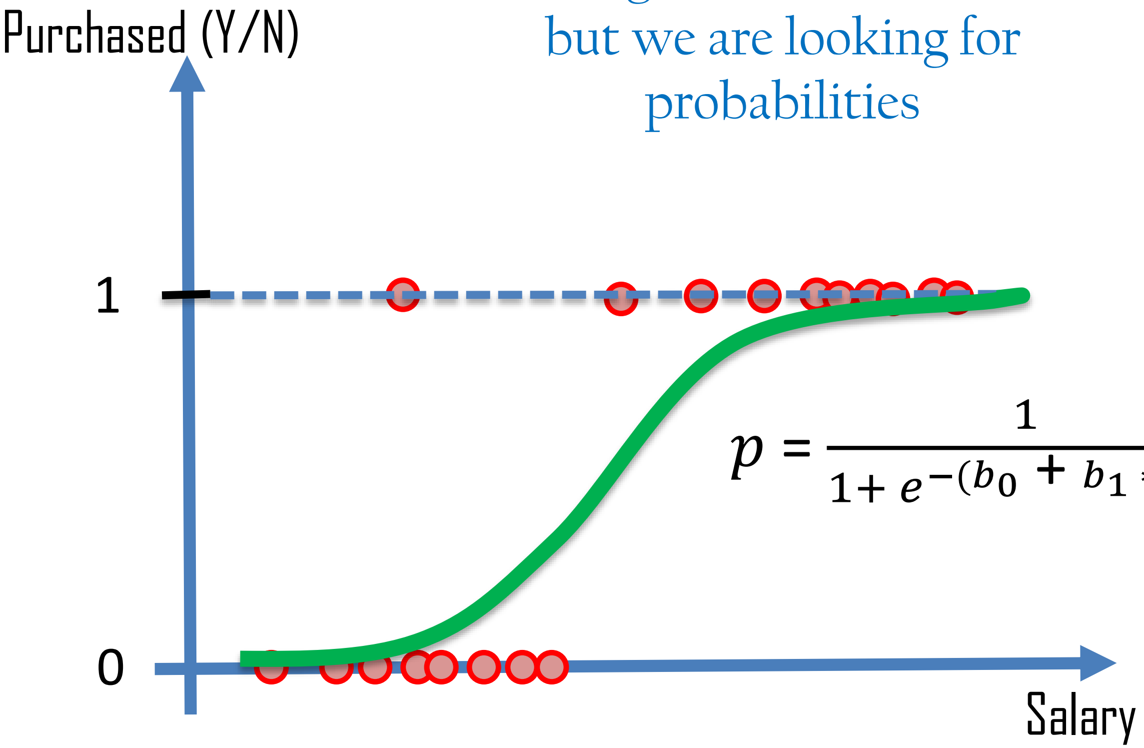
$$\log_e\left(\frac{P}{1-P}\right) = b_0 + b_1 * x_1$$

$$p = \frac{1}{1+e^{-(b_0 + b_1 * x_1)}}$$



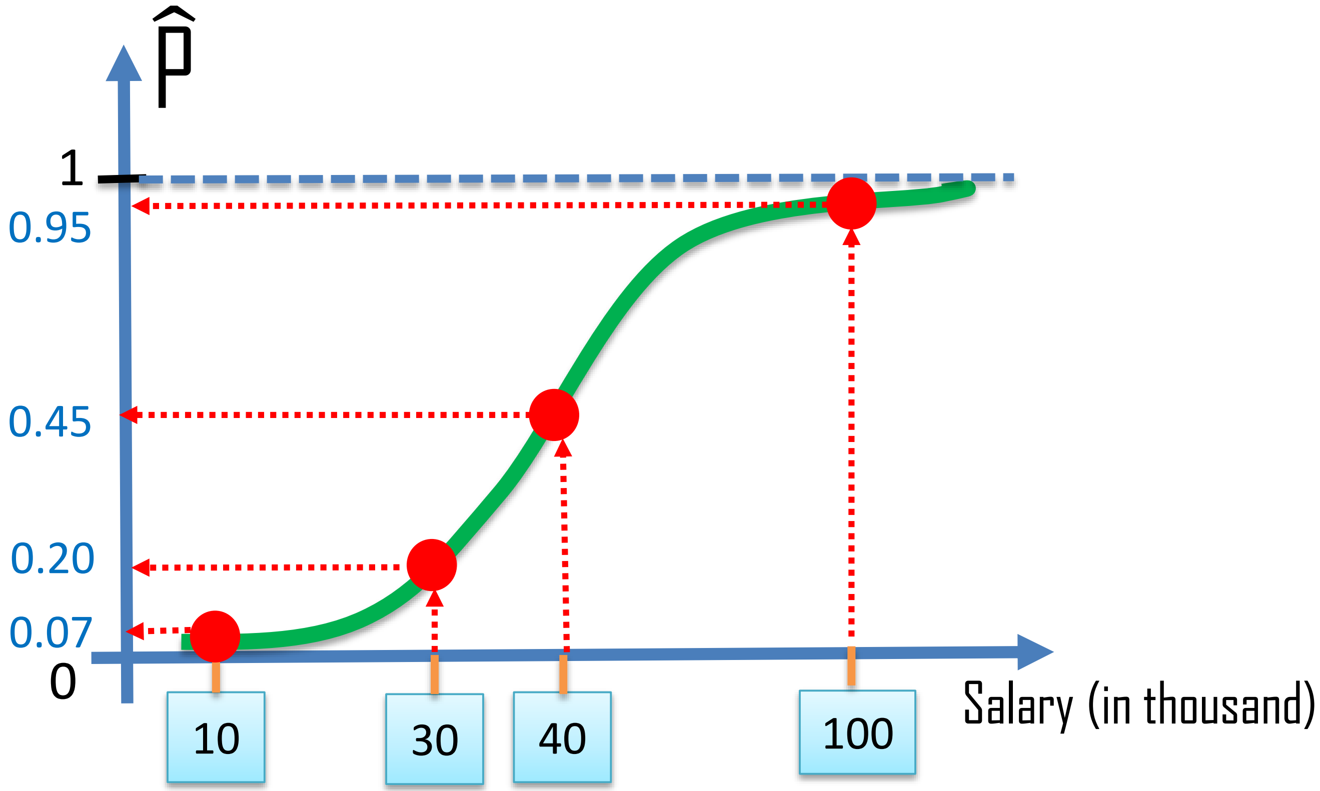
Logistic Regression

Logistic regression is same as linear regression as best fit line but we are looking for probabilities

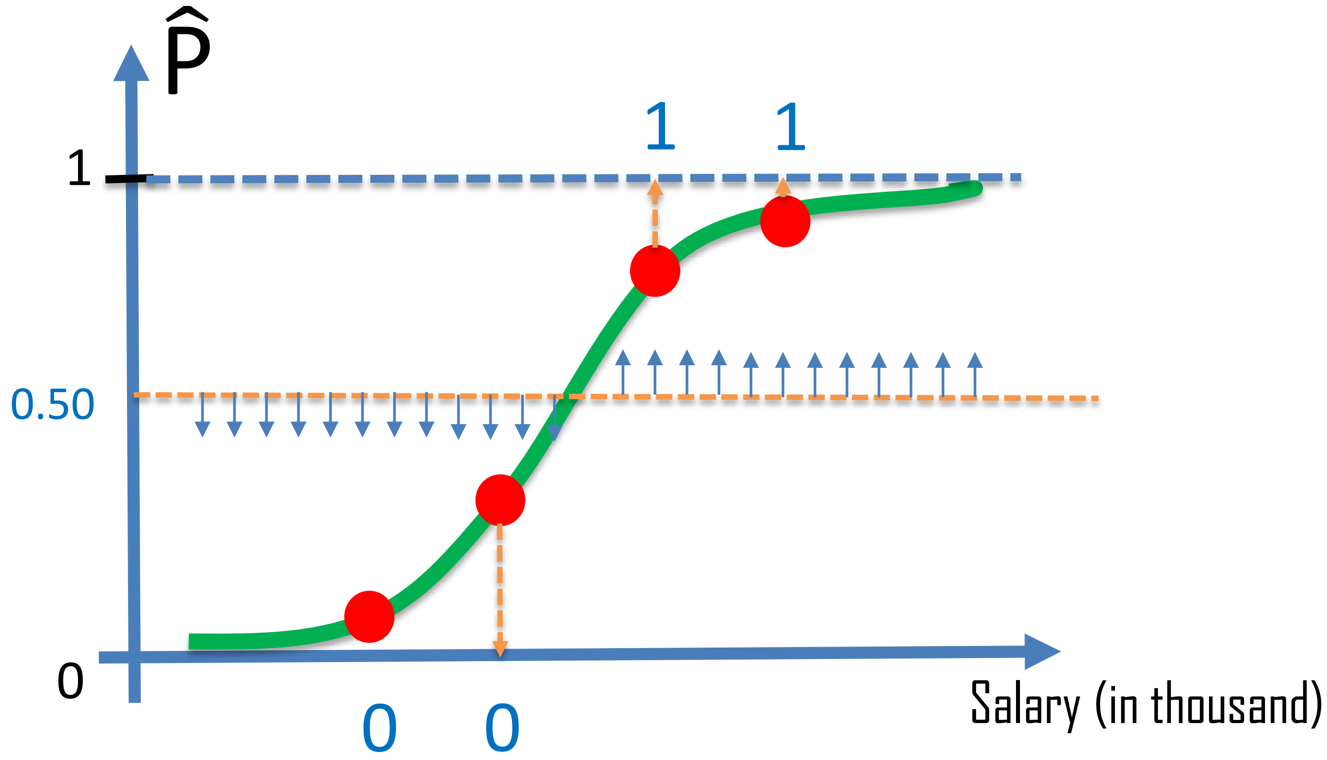


$$p = \frac{1}{1+ e^{-(b_0 + b_1 * x_1)}}$$

Logistic Regression



Logistic Regression



Classification Model Performance

True Negatives (TN): Actual FALSE, which was predicted as FALSE

False Positives (FP): Actual FALSE, which was predicted as TRUE (Type I error)

False Negatives (FN): Actual TRUE, which was predicted as FALSE (Type II error)

True Positives (TP): Actual TRUE, which was predicted as TRUE

Classification Performance Metric

Metric	Description	Formula
Accuracy	What % of the prediction were correct?	(TP + TN)/ (TP + TN + FP + FN)
Misclassification rate	What % of prediction were wrong ?	(FP + FN)/ (TP + TN + FP + FN)
True Positive rate or Sensitivity or Recall	What % of positive classes did model catch ?	TP / (TP + FN)
False positive rate	What % of "No" were predicted "Yes"	FP / (FP + TN)
Specificity	What % of "No" were predicted "No"	TN / (TN + FP)
Precision (exactness)	what % of positive predictions were correct?	TP/(TP+FP)
F1 score	Weighted average of precision and recall	2*((precision * recall) / (precision + recall))

INNOMATICS TECHNOLOGY HUB

Hands on Logistic Regression

INNOMATICS

TECHNOLOGY HUB