Logistic Regression

Friday, 23 December 2022 12:22 AM

Regression

Continuous Columns

Classification

Categorical Columns

Mathematical Intuition

$\hat{y} = \beta + \beta_1 x\_1$

$-\infty \; to \; \infty$

$-\infty \; to \; \infty$

But for Logistic Regression

$\hat{y} \neq \beta + \beta_1 x_1$

This comes as Probability

$0 \; to \; 1$

To overcome this

$P = \beta + \beta_1 x_1$

$\frac{P}{1-P} \neq \beta + \beta_1 x_1$

$0 \; to \; \infty$

Odds: Probability of getting something upon Probability of not Finding something

To overcome this

$\log \frac{P}{1-P} = \beta + \beta_1 x_1$

$-\infty \; to \; \infty$

Solving this

$\log \frac{P}{1-P} =$

$\frac{P}{1-P} = e^{\beta}$

$P = e^{\beta+\beta x}$

$P(1+e^{\beta+\beta x}$

$P = \frac{e^{\beta+\beta x}}{1+e^{\beta+\beta x}}$

$\rightarrow \quad -\infty\ to\ \infty$

$;$

$= \beta + \beta x$

$+ \beta x$

$- P * e^{\beta+\beta x}$

$) = e^{\beta+\beta x}$

Stats Model

Linear Regression

Logistic Regressi

Interpretation is possible because stats model can give you sum

$$P = \frac{e^{\beta + \beta x}}{1 + e^{\beta + \beta x}}$$

We always get Probability values, after this we cla
using a P threshold

Classification

Binary Classification

Multiclass Class

Binary Classification: We have only 2 sub categories in dependent column
Multiclass Classification: We have more than 2 subcategories in dependent column

Model Performance Methods

- Performance can be calculated by comparing 2 models

Maximum Log Likelihood

$$LL = \sum_{i=1}^{N} y \log P(y) + (1 - y) \log(1 - P(y))$$

- Gives Performance of model
- Higher is better
- It is boundless
- Value always in negative

Deviance

$$Deviance = -2 * LL$$

- Lower is better

Two Types of Deviance:

Null Deviance: When not a single i

on

mary

assify it into a category

sification

ndependent

- Value always in negative

Log Loss Function

$$Log\ Loss\ = \frac{-1}{n} * LL$$

- *Always in positive*
- *Lower is better*

Bayesian Information Criteria

$$BIC\ = -2LL + k\log n$$

- K is total no features
- N is total no rows
- It accounts both rows and features to calculate model perfomance

Feature is considered while calculat

Residual Deviance: When all indep
Are considered while calculating LL

Akaike Information Criteria/AIC

$$AIC\ = -2LL + 2k$$

- K is total no of features
- It accounts features in calcula
performance
- Lower is better

Model Evaluation

Confusion Matrix

Predicted

Actual

	False	True
False	**TN**	**FP**
True	**FN**	**TP**

Confusion Matrix T

Actual

Predicted

	False
False	**TN**
True	**FP**

Given Errors:

Category	**Actual**	**Predicted**
Accepted	450	600
Rejected	550	400

Confusion Matrix ?

Predicted

Confusion Matrix Transp

Actual

	Rejected	A
Rejected	**T400**	
Accepted	**F0**	

ting LL

pendent features

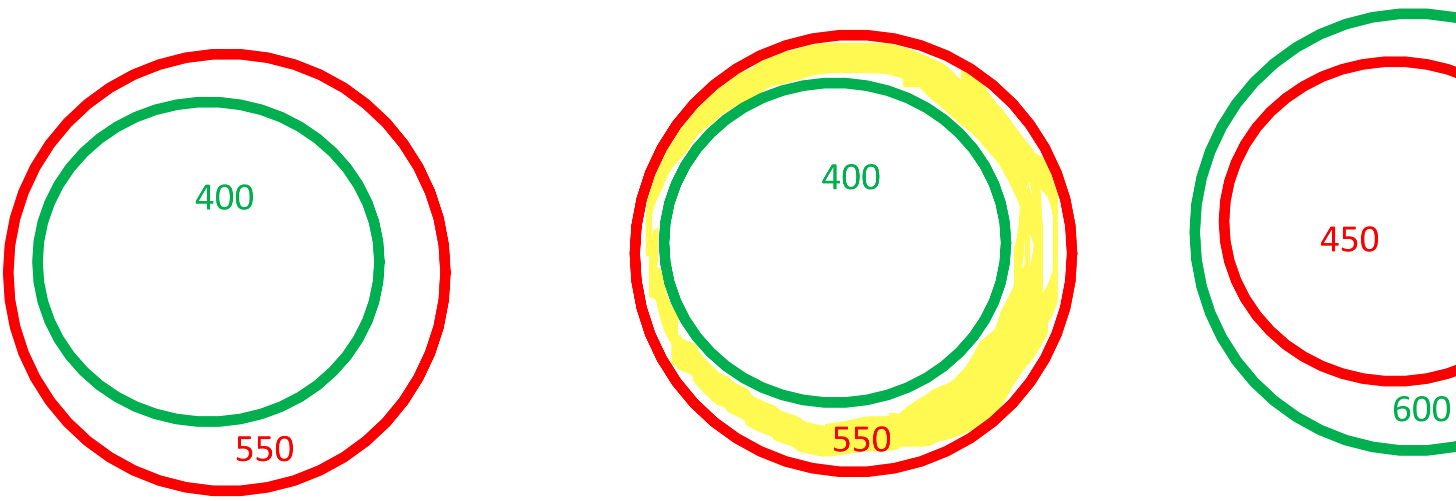
ating model

<u>Transposed</u>

True
FN
TP

<u>posed</u>

ccepted
F150
T450



Accuracy

$$Accuracy = \frac{Total\ True\ Prediction}{Total\ Values} = \frac{TP + TN}{TP + FN + FP + TN}$$

- Higher is better
- In percentage
- Chances are that model identifies TP but not TN, in such case

Precision

$$Precision = \frac{True\ Positive}{Total\ Predicted\ Positive\ Values} = \frac{TP}{TP + FP}$$

Recall

$$Recall = \frac{True\ Positive}{Total\ Actual\ Values} = \frac{TP}{TP + FN}$$

F1 SCORE

$$F1 = \frac{-2PR}{P + R}$$

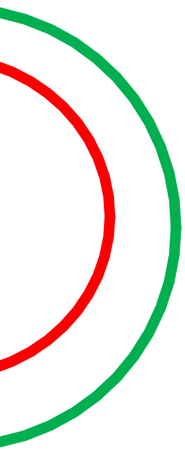
- It is the harmonic
- Higher is better

Probability Threshold

The value used to determine which class observed value should be classified as, Changing can render all above calculation as useless, To over come this

True Prediction Rate/TPR

$True\ Postive\ Prediction \qquad TP$



c mean of precision & recall

this threshold

↑ TPR

$$TPR = \frac{True\ Positive\ Prediction}{Total\ Actual\ Positive} = \frac{TP}{(TP + FN)}$$

FPR $\downarrow$

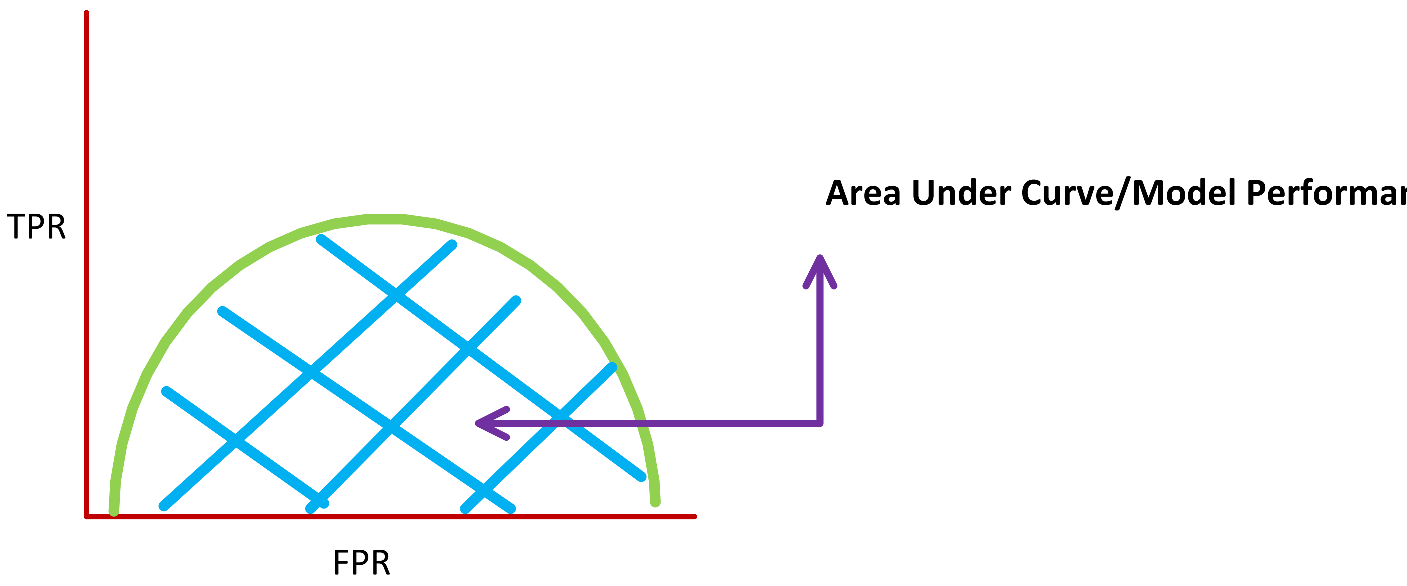
False Prediction Rate

$$FPR = \frac{False\ Positive\ Prediction}{Total\ Actual\ Negative\ Values} = \frac{FP}{FP + TN}$$

Our main objective is to decrease FPR, Increase TPR

Receiver Operating Characteristics/Area under Curve

- When TPR plotted on y and FPR on x, Area formed under curve will evaluate model independent of threshold
- It can also be used to figure the best Probability Threshold for model



performance

nce