Supervised Learning Classification

Logistic Regression

Agenda

- Introduction to Classification
- Introduction to Logistic Regression
- Difference between Regression and Classification
- Assumption of Logistic Regression
- Types of Logistic Regression
- Implementation of Logistic Regression / Working
- Model Evaluation in Classification

Introduction to Classification

Classification is technique where we categorize data into given number of classes.

The main goal of classification problem is identify the category / class to which a new data will fall under.

1) Binary classification :- task with two possible outcome.(T / F)

2) Multi-class classification :- more than two classes.

3) Multi-labels classification :- Classification task where each sample is mapped to a set of target labels. Ex :- A news article can be about sport, a person, etc.

Introduction to Classification

Logistic regression used when the target variable is categorical. Hence it can be used for classification:

- Commonly used to estimate the probability that an instance belong to a particular class.
- If estimate probability is greater than 50% then the model predict that the instance belong to that class. (labeled '1' or called the '+ve' class) or else predict that the is does not (i.e. labeled '0' or called the '-ve' class).
- Our value of Y will be between 0 and 1.
- Logistic regression is named for the function used at the core of the method the " Logistic Function" also called as " Sigmoid Function".

Difference between Classification & Regressions

CLASSIFICATION	REGRESSION
Is the task of predicting a discrete class label.	Is the task of predicting a continuous quantity
Prediction can be evaluated using accuracy.	Prediction can not be evaluated using accuracy. MSE, RMSE, etc.
Can have real-values or discrete value.	Real value such as integer or floating value.
More than two classes is often called multi-class classification problem.	A problem with multiple i/p variables is often called a multivariate regression problem.

Linear Regression vs Logistic Regression

LINEAR REGRESSION	LOGISTIC REGRESSION
Continuous variable	Categorical or discrete variable
Solve regression problem	Solve classification problem.
Straight line.	S-curve

Logistic Regression

- The logistic function, also called the sigmoid function.
- It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$$1 / (1 + e^{-value})$$

- Where, e is the base of the natural logarithms (Euler's number or EXP() number).
- Transformed into the range 0 and 1 using the logistic function.

Logistic Regression

- Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y).
- A key difference from linear regression is that the output value being modeled is a binary values (0 or 1) rather than a numeric value.

Logistic Regression

- The equation is:

$$y = e^{(b0 + b1*x)} / (1 + e^{(b0 + b1*x)})$$

- Where,
    - y is the predicted output,
    - b0 is the bias or intercept term and b1 is the coefficient for the single input value (x).
    - Each column in your input data has an associated b coefficient (a constant real value) that must be learned from your training data.

Logistic Regression

For example, if we are modeling people's sex as male or female from their height, then the first class could be male and the logistic regression model could be written as the probability of male given a person's height, or more formally:

$$P(\text{ sex} = \text{male} \mid \text{height })$$

Logistic Regression

- Note that the probability prediction must be transformed into a binary values (0 or 1) in order to actually make a probability prediction.
- Logistic regression is a linear method, but the predictions are transformed using the logistic function.
- for example, continuing on from above, the model can be stated as:

$$p(X) = e^{(b0 + b1*X)} / (1 + e^{(b0 + b1*X)})$$

- The above equation as follows (remember we can remove the e from one side by adding a natural logarithm (ln) to the other):

$$\ln(p(X) / 1 - p(X)) = b0 + b1 * X$$

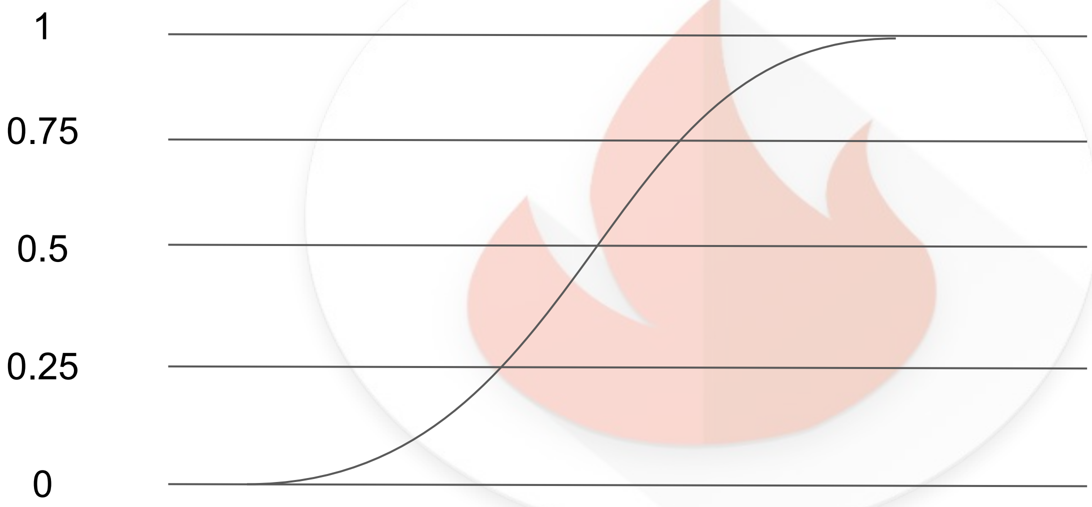
Logistic Regression Predict Probabilities

- This is useful because we can see that the calculation of the output on the right is linear again (just like linear regression), and the input on the left is a log of the probability of the default class.
- This ratio on the left is called the odds.
- Odds are calculated as a ratio of the probability of the event divided by the probability of not the event, e.g. 0.8/(1-0.8) which has the odds of 4.
- So we could instead write: ln(odds) = b0 + b1 * X
- Because the odds are log transformed, we call this left hand side the log-odds or the probit.

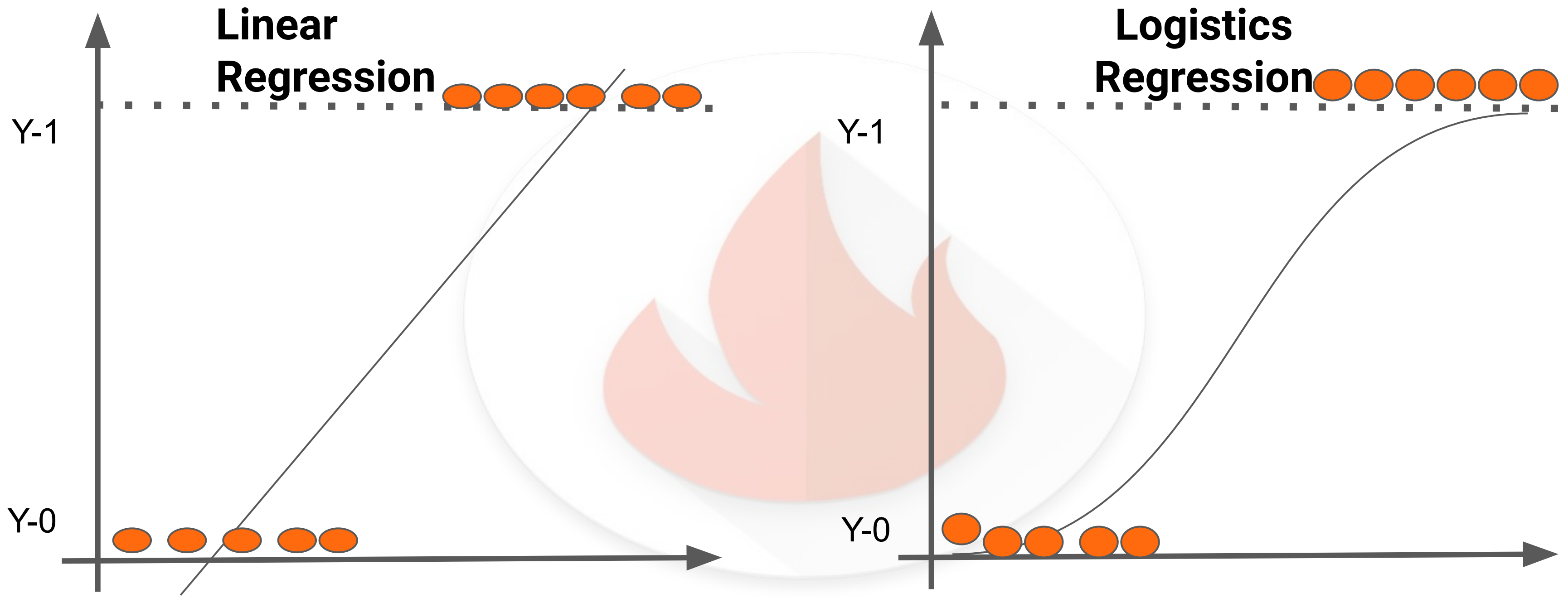
Logistic Regression Predict Probabilities

- Because the odds are log transformed, we call this left hand side the log-odds or the probit.
- We can move the exponent back to the right. odds = e^(b0 + b1*X)

Logit Function



Comparison of Linear and Logistic Regression



Learning the Logistic Regression Model

- The coefficients (beta) of the logistic regression algorithm must be estimated from your training data.
- This is done by using maximum-likelihood estimation. (MLE)
- The best coefficients would result in a model that would predict a value very close to 1 (belongs to class for ex.male)and a value very close to 0 (belongs to class for ex.female).
- The MLE that minimize the error in the probabilities predicted by the model. MLE is a minimization algorithm is used to optimize the best value for the coefficients for your training data.

Assumptions of Logistic Regression

- **Binary output variable :-** This might be obvious. Logistic regression is intended for binary (two class) classification problem.
- **Remove Noise :-** Logistic regression assumes no error in the o/p variable (Y).
- **Gaussian Distribution :-** Logistic Regression is a linear regression. It does assume a linear relationship between the i/p variables with the o/p.
- **Remove correlated i/p :-** The model can overfit if you have multiple highly-correlated i/p. Consider calculating the pairwise correlation between all i/p & removing highly correlated i/p.
- **Fail to Converge :-** This can happen if there are many highly correlated i/p in dataset.(e.g. lots of zeros in your input data).

Types of Logistic Regression

- **Binary Logistic Regression :-** The categorical response has only two possible outcomes. Example :- email spam or not (hamp)
- **Multinomial Logistic Regression :-** Three or more categorical without ordering. Example :- Which food is preferred more(veg, non-veg, vegan.)
- **Ordinal Logistic Regression :-** Three or more than categories with ordering. Example :- Movie rating from 1 to 5.

Model Evaluation in Classification

- Classification accuracy is the ratio of corrected predictions to total predictions mode.
- Classification accuracy = correct prediction / total predictions
- It is often presented as a percentage by multiplying 100 But, classification accuracy is that it hides the detail you need to better understand the performance of your classification model

Model Evaluation in Classification

A confusion matrix is a summary of prediction results on a classification problem.

The number of correct and incorrect predictions are summarized with count values and broken down by each class.

<u>Below is the process for calculating a confusion matrix.</u>

1. You need a test dataset.
2. Make a prediction for each row in your test dataset.
3. From the expected outcomes and prediction count:
   - The number of correct predictions for each class
   - The number of incorrect predictions for each class, organized by class that was predicted.

Model Evaluation in Classification

- These number are then organized into a table, or a matrix:
  - Each row of the matrix corresponds to a predicted class.
  - Each column of the matrix corresponds to an actual class.

	Actual Values	
Predicted Values	TP	FP
	FN	TN

Model Evaluation in Classification

Expected	Predicted
man	woman
man	man
woman	woman
man	man
woman	man
woman	woman
woman	woman
man	man
man	woman
woman	woman

Model Evaluation in Classification

- Total number of correct predictions for a class go into the expected row for that class value and the predicted column for that class value.
- accuracy = total correct predictions / total predictions made * 100

accuracy = 7 / 10 * 100

- Let's turn our results into a confusion matrix.
- First,we must calculate the number of correct predictions for each class.
  - man classified as man: 3
  - woman classified as woman : 4

Model Evaluation in Classification

- Now, we calculate the number of incorrect predictions for each class, organized by the predicted value.
  - man classified as woman: 2
  - woman classified as man : 1

Now, arrange these values into the 2-class confusion matrix

	man	woman
man	3	1
woman	2	4

- The correct values are organized in a diagonal line (3+4)

Model Evaluation in Classification

- In this way, we can assign the event row as “positive” and the no-event row as “negative”.
- We can then assign the event predictions as ‘T’ and the no-event as ‘F’.
- Now, arrange these values into the 2-class confusion matrix

	event	no-event
event	TP	FP
no-event	FN	TN

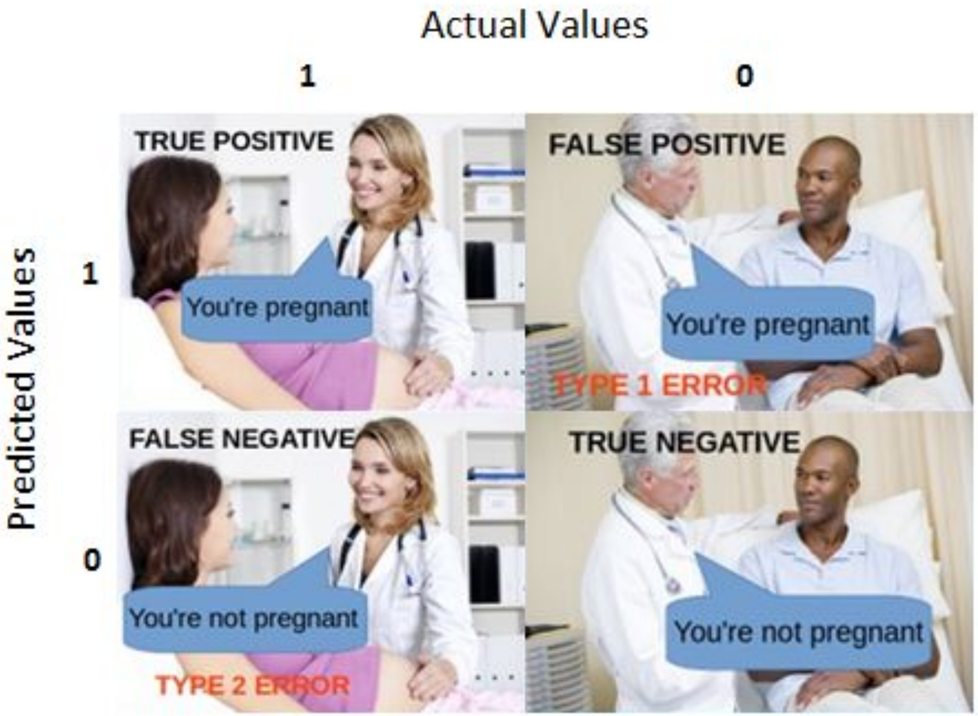
Model Evaluation in Classification

This gives us:

- **'True Positive'**: for correctly predicted event values.
- **'False Positive'**: for incorrectly predicted event values.
- **'True Negative'**: for correctly predicted no-event values.
- **'False Negative'**: for incorrectly predicted no-event values.

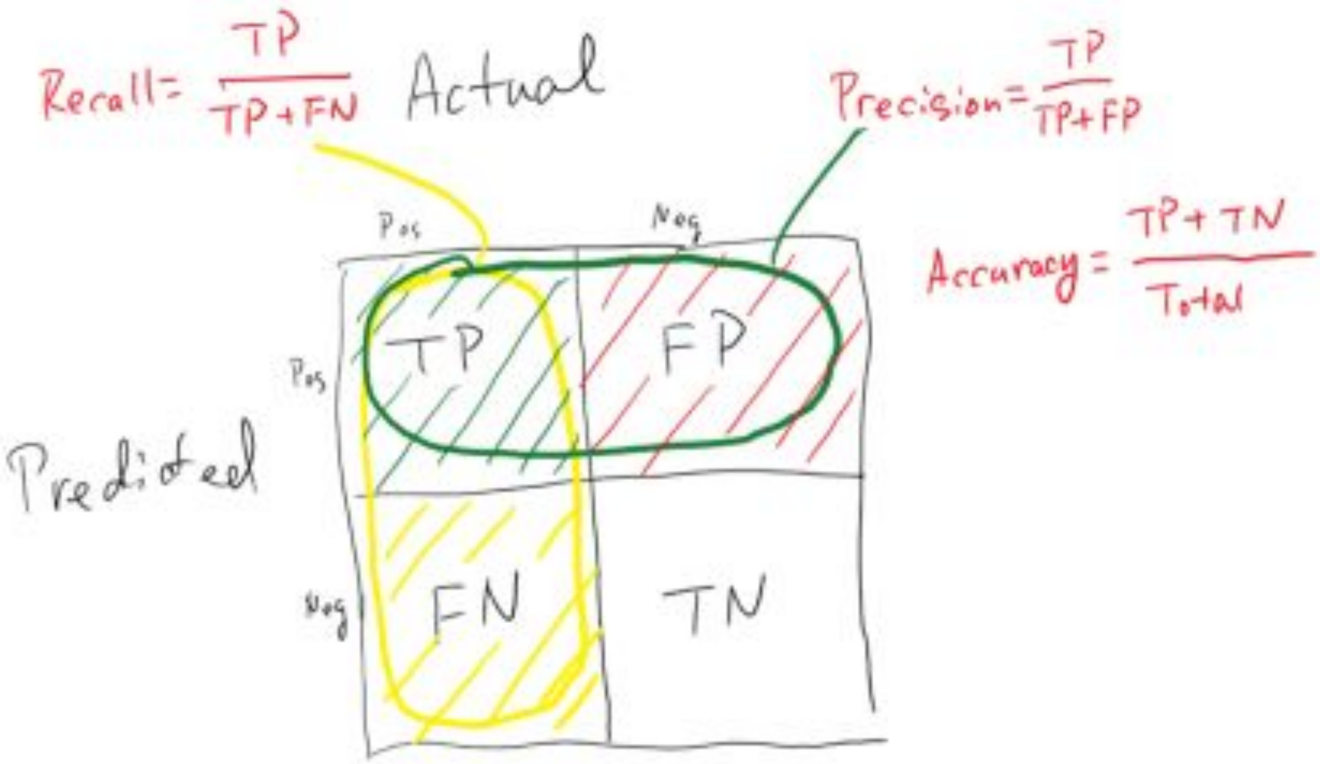
It is extremely useful for measuring Recall, Precision, Specificity, Accuracy, and most useful AUC-ROC curve.

Model Evaluation in Classification



Picture Source :- Google

Model Evaluation in Classification



Picture Source :- Google

Model Evaluation in Classification

F-Measure provides a way to combine both precision and recall into a single measure that captures both properties.

$$F\text{-}measure = 2 * (Recall * Precision) / (Recall + Precision)$$

Thank you