UEM
श्रद्धावान लभते ज्ञानम्
UNIVERSITY OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

Course Name : AI & ML

Logistic Regression?

It’s a classification algorithm.

Target value is categorical.

When to use?

logistic regression should only be used when the target variables fall into discrete categories and that if there’s a range of continuous values the target value might be predicted, logistic regression should not be used.

Binary Classification Problems

$y \in \{0, 1\}$

where 0: Negative class

1: Positive class

Based on some threshold value problems are being classified.

How it differs from Linear Regression?

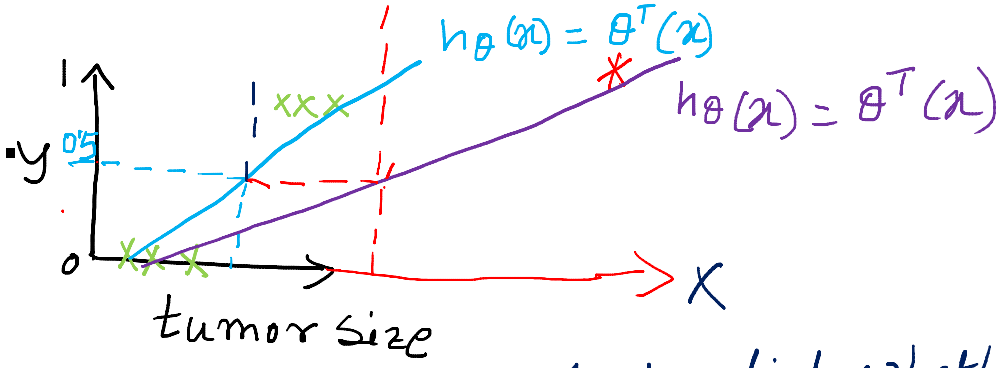
Linear regression is not appropriate in case of a qualitative response.

Can we apply Linear Regression to classification??

How Linear Regression Works?

It generally takes some ‘x’ value to predict a ‘h(x)’, provided there is some parameter (regression coefficient) exists.

It produce a straight line, originating from a starting (x, y) towards an ending (x, y). Each point over the line is a predicted value.



$h_\theta(x) \geq 0.5, \; y = 1$

$h_\theta(x) < 0.5, \; y = 0$

Here, we are trying to predict whether a tumor is malignant or not based on its size. We have tried to apply linear regression $h_\theta(x) = \theta^T(x)$. In the 1st case (blue line), regression line is being drawn through present points (green). If we check the '0·5' level and its impact on 'x' axis, then we can see that all the tumors which are predicted as '1' are in the right side of the blue dotted line. It indicates that linear regression can work fine. But, in the next case (purple line) when one very large tumor is present, regression line is being changed and impact of '0·5' level over 'x' axis. It makes previous +ve class (1) into -ve class. So, linear regression is not good for classification.

Other than the above explanations, another reason for not using linear regression is the value. Classification demands the value as 0 or 1. But regression equation may result in any value.

Solution comes with logit function which can output values in the range of 0 to 1. From that value with a predefined assumption classification can be made.

Therefore, logistic regression is expressed as:

$0 <= h\theta(x) <= 1$

So, that classification can be done properly.

Sigmoid Function:

hθ(x)=g(Θ^T x).....(i)

g(z)= 1/ 1+ e ^(-z).....(ii) [z is a real number]

This function is called sigmoid function or logistic function.

By combining (i) & (ii)....

hθ(x)=1/1+e^(-(Θ^T x)) (if we draw a graph we can understand the significance of the hypothesis value which is b/w 0 to 1)

- How sigmoid function looks?

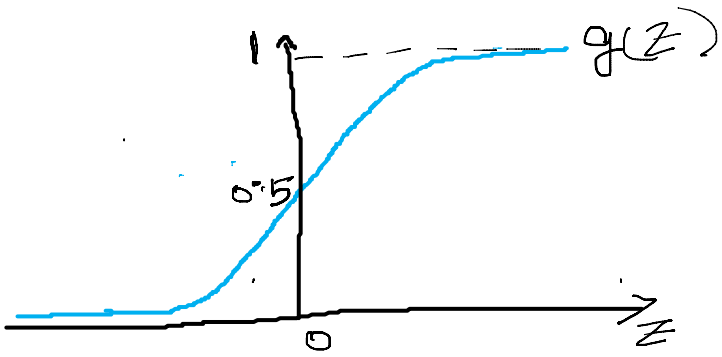
In logistic regression –

$h_\theta(x) = g(z)$

where, $g(z) = \frac{1}{1+e^{-z}}$

as $h_\theta(x) = \theta^T(x)$

so, $z = \theta^T(x)$



y [redacted] will be predicted as 1, if $z \geq 0$

else y will be predicted as 0.

A group of 20 customers, having yearly savings between 0 and 6 lacs, are considered for analysis. We need to check how does savings of a customer is related to repayment of loan? [$\beta 0 = -4.07778$, $\beta 1 = 1.5046$]

Savings (lacs)	Loan Non-Defaulter	Savings (lacs)	Loan Non-Defaulter
0.5	0	2.75	1
0.75	0	3.00	0
1.00	0	3.25	1
1.25	0	3.50	0
1.5	0	4.00	1
1.75	0	4.25	1
1.75	1	4.5	1
2	0	4.75	1
2.25	1	5.00	1
2.5	0	5.5	1

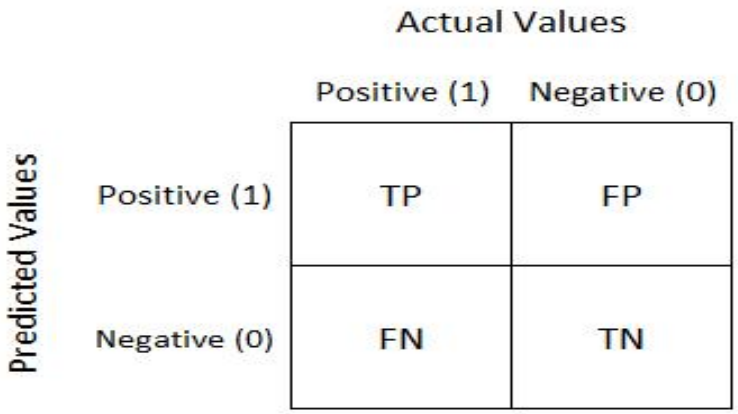
Apply sigmoid function to estimate the outcome:
0 : defaulters, 1 : non defaulters

Loan Non-Defaulter	Predicted	Loan Non-Defaulter	Predicted
0	0	1	1
0	0	0	1
0	0	1	1
0	0	0	1
0	0	1	1
0	0	1	1
1	0	1	1
0	0	1	1
1	0	1	1
0	0	1	1

Confusion Matrix:-

A confusion matrix is a table that is often used to **describe the performance of a classification model** (or "classifier") on a set of test data for which the true values are known.

Summary



1.Accuracy (all **correct** / all) = TP + TN / TP + TN + FP + FN
2.Misclassification (all **incorrect** / all) = FP + FN / TP + TN + FP + FN
3.Precision (**true** positives / **predicted** positives) = TP / TP + FP
4.Sensitivity aka Recall (**true** positives / all **actual** positives) = TP / TP + FN
5.Specificity (**true** negatives / all **actual** negatives) =TN / TN + FP

- **true positives (TP):** These are cases in which we predicted yes (they have the disease), and they do have the disease.
- **true negatives (TN):** We predicted no, and they don't have the disease.
- **false positives (FP):** We predicted yes, but they don't actually have the disease. (Also known as a "Type I error.")
- **false negatives (FN):** We predicted no, but they actually do have the disease. (Also known as a "Type II error.")

List of rates that are often computed from a confusion matrix for a binary classifier:

- **Accuracy:** Overall, how often is the classifier correct?
  - (TP+TN)/total = (100+50)/165 = 0.91
- **Misclassification Rate:** Overall, how often is it wrong?
  - (FP+FN)/total = (10+5)/165 = 0.09
  - equivalent to 1 minus Accuracy
  - also known as "Error Rate"
- **True Positive Rate:** When it's actually yes, how often does it predict yes?
  - TP/actual yes = 100/105 = 0.95
  - also known as "Sensitivity" or "Recall"
- **False Positive Rate:** When it's actually no, how often does it predict yes?
  - FP/actual no = 10/60 = 0.17
- **True Negative Rate:** When it's actually no, how often does it predict no?
  - TN/actual no = 50/60 = 0.83
  - equivalent to 1 minus False Positive Rate
  - also known as "Specificity"
- **Precision:** When it predicts yes, how often is it correct?
  - TP/predicted yes = 100/110 = 0.91
- **Prevalence:** How often does the yes condition actually occur in our sample?
  - actual yes/total = 105/165 = 0.64

Problems on confusion matrix

- We have a total of 2**0 cats and dogs** and our model predicts whether it is a cat or not. Calculate performance of the model.
- **Actual values = [‘dog’, ‘cat’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘cat’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘dog’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘cat’]**
  Predicted values = [‘dog’, ‘dog’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘cat’, ‘cat’, ‘cat’, ‘cat’, ‘dog’, ‘dog’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘cat’, ‘dog’, ‘dog’, ‘cat’]

Problems on confusion matrix

- The classifier made a total of 165 predictions (e.g., 165 patients were being tested for the presence of that disease).
- Out of those 165 cases, the classifier predicted "yes" 110 times, and "no" 55 times.
- In reality, 105 patients in the sample have the disease, and 60 patients do not.

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Odds & Odd Ratio

Odds=(probability of the event)/(probability of the non-event)

- Odds of an event happening is defined as the likelihood that an event will occur, expressed as a proportion of the likelihood that the event will not occur. Therefore, if A is the probability of subjects affected and B is the probability of subjects not affected, then odds = A /B.
- Therefore, the odds of rolling four on a dice are 1/5 or 20%.
- **Odds Ratio** (OR) is a measure of association between exposure and an outcome. The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure.

- **Important points about Odds ratio:**
- Calculated in case-control studies as the incidence of outcome is not known
- OR >1 indicates increased occurrence of an event
- OR <1 indicates decreased occurrence of an event

Example

- Probabilities range between 0 and 1. Let's say that the probability of success is .8, thus
- **p = .8**
- Then the probability of failure is
- **q = 1 – p = .2**
- Odds are determined from probabilities and range between 0 and infinity. Odds are defined as the ratio of the probability of success and the probability of failure. The odds of success are
- **odds(success) = p/(1-p) or p/q = .8/.2 = 4,**
- that is, the odds of success are 4 to 1. The odds of failure would be
- **odds(failure) = q/p = .2/.8 = .25.**

Decision Boundary

Decisions are being taken based on the value of hθ(x). If it is less than 0.5 then ‘y’ is 0 else ‘y’ value is 1.

Now if hθ(x) = g(Θ0 + Θ1 x1 + Θ2 x2), then how decisions will be taken?

‘y’ will be 1, if z>=0, now ‘z’ means Θ^T x, so Θ^T x >= 0.

Based on the value of parameters we will get a linear equation which is basically a straight line. This line will create a proper boundary between positive & negative cases.

Decision Boundary

Now if $h_\theta(x) = g(\theta 0 + \theta 1\, x1 + \theta 2\, x2)$, then how decisions will be taken?

Here, $h_\theta(x)$ is giving a straight line. We need to assume some values for $\theta_0, \theta_1, \theta_2$ for which $z \geq 0$ will satisfy.
z means $\theta^T x$. To predict $y=1$, z should be $z \geq 0$.
Say $\theta_0 = -3$, $\theta_1 = 1$, $\theta_2 = 1$. So, $-3 + x_1 + x_2 \geq 0$
or $x_1 + x_2 = 3$

Non linear decision boundary

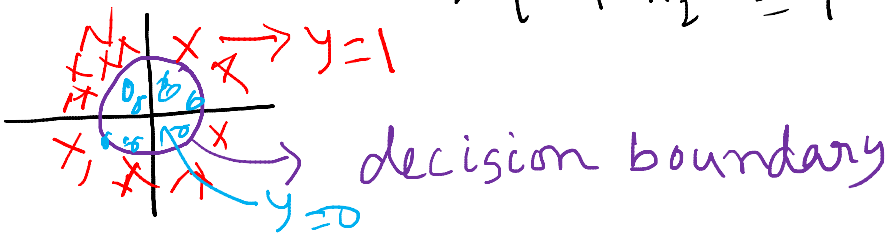
Ex: $h\theta(x) = g(\theta 0 + \theta 1\, x1 + \theta 2\, x2 + \theta 3\, (x1)^2 + \theta 4\, (x2)^2)$?

Here, polynomial term indicates that it can not be straight line. Now, say $\theta_0 = -1,\ \theta_1 = 0,\ \theta_2 = 0,\ \theta_3 = 1,\ \theta_4 = 1$.

Then, $-1 + 0 \cdot x_1 + 0 \cdot x_2 + x_1^2 + x_2^2 \geq 0$ [As $z \geq 0$]

or $x_1^2 + x_2^2 \geq 1$

$\simeq$ $x_1^2 + x_2^2 = 1 \longrightarrow$ A circle equ:



Cost function of linear regression involves square terms. As a result it produce a non convex output because of non-linear term of the hΘ(x). From a non convex function its not possible to obtain a global minimum using gradient descent. So, different set of equations are used

Cost function of logistic regression is not same as linear regression.

Cost ($h_\theta(x)$, y) = -log($h_\theta(x)$) predicts y=1

= -log(1 - $h_\theta(x)$) predicts y=0

This equation produces a graph which indicates when $h_\theta(x) = 1$ & y=1 then cost 0.

But, if y=1 & $h_\theta(x) = 0$, cost will be infinite.

For y =0 situation is just vice-versa.

Simplified equation:

Cost ($h_\theta(x)$, y)= -y log($h_\theta(x)$) – (1-y) log(1-$h_\theta(x)$)

Multi-class Classification:

More than 2 class.

One-vs-All classification: Every time use the binary classification concept to segregate one class from the rest. Repeats the same till all classifications are done.

Thank You