

Logistics Regression

Classification

We want to learn about Logistic Regression as a method for **Classification**.

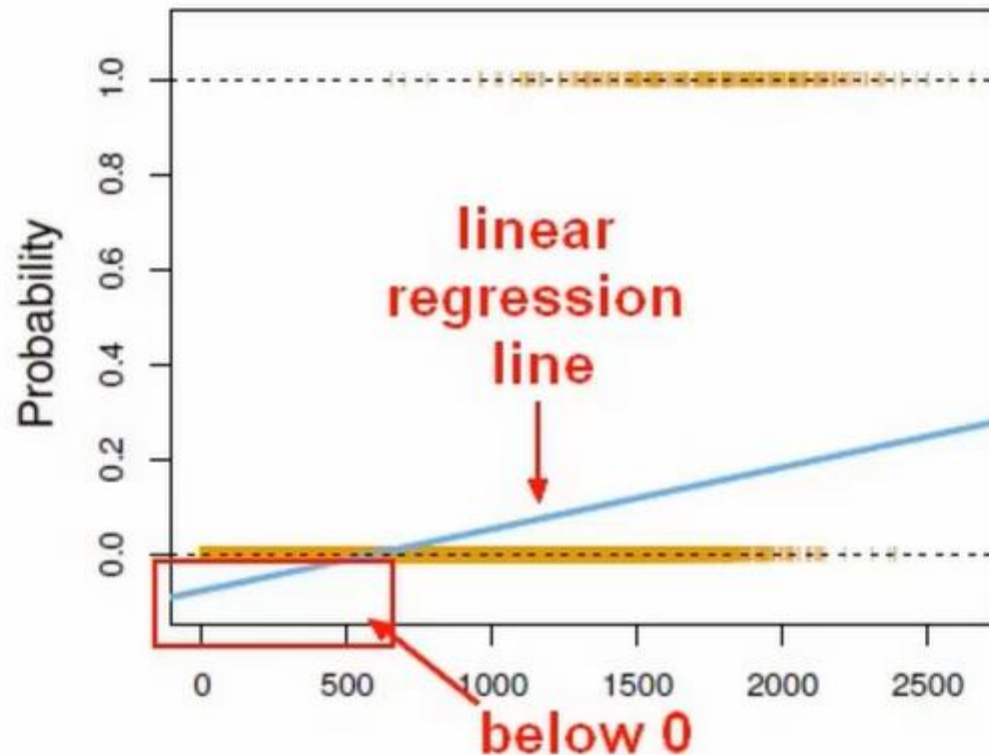
Some examples of classification problems:

- Spam versus “Ham” emails
- Loan Default (yes/no)
- Disease Diagnosis

Above were all examples of Binary Classification

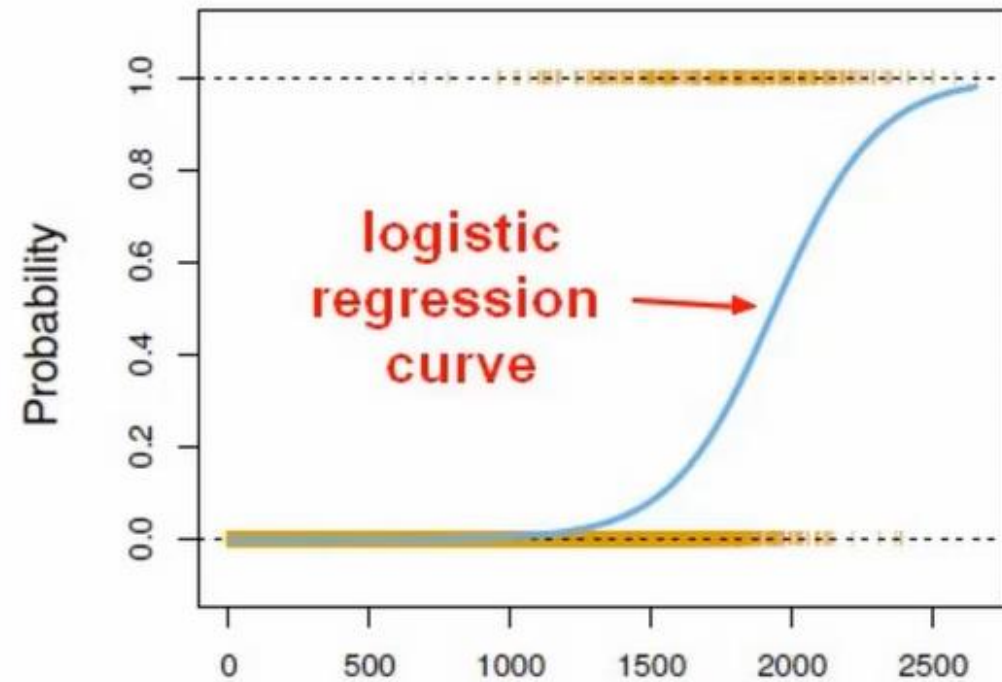
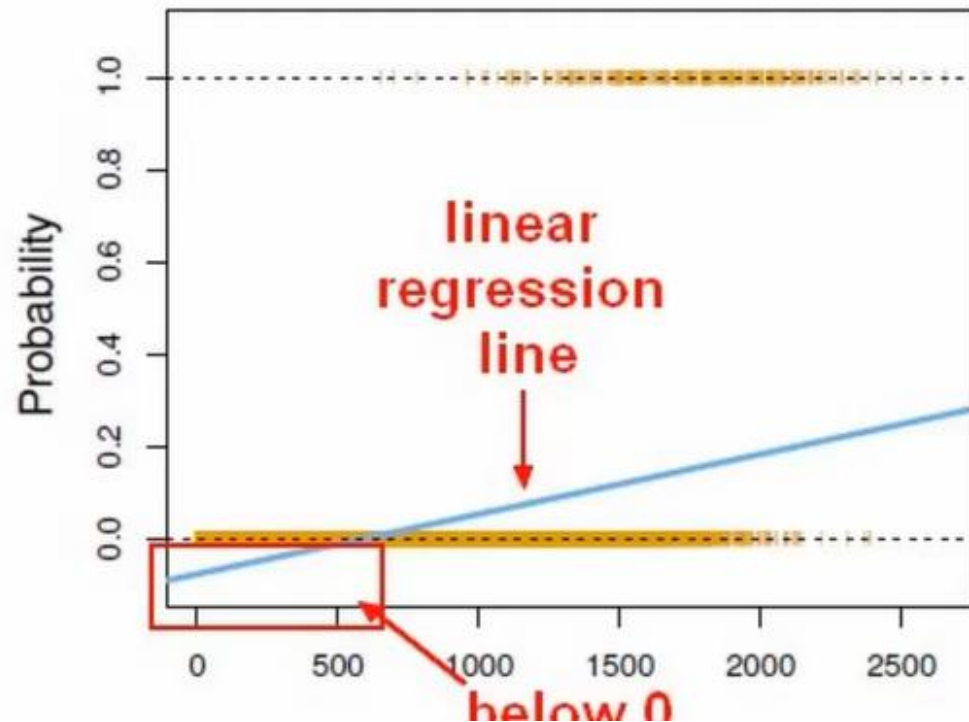
Classification

We can't use a normal linear regression model on binary groups. It won't lead to a good fit:



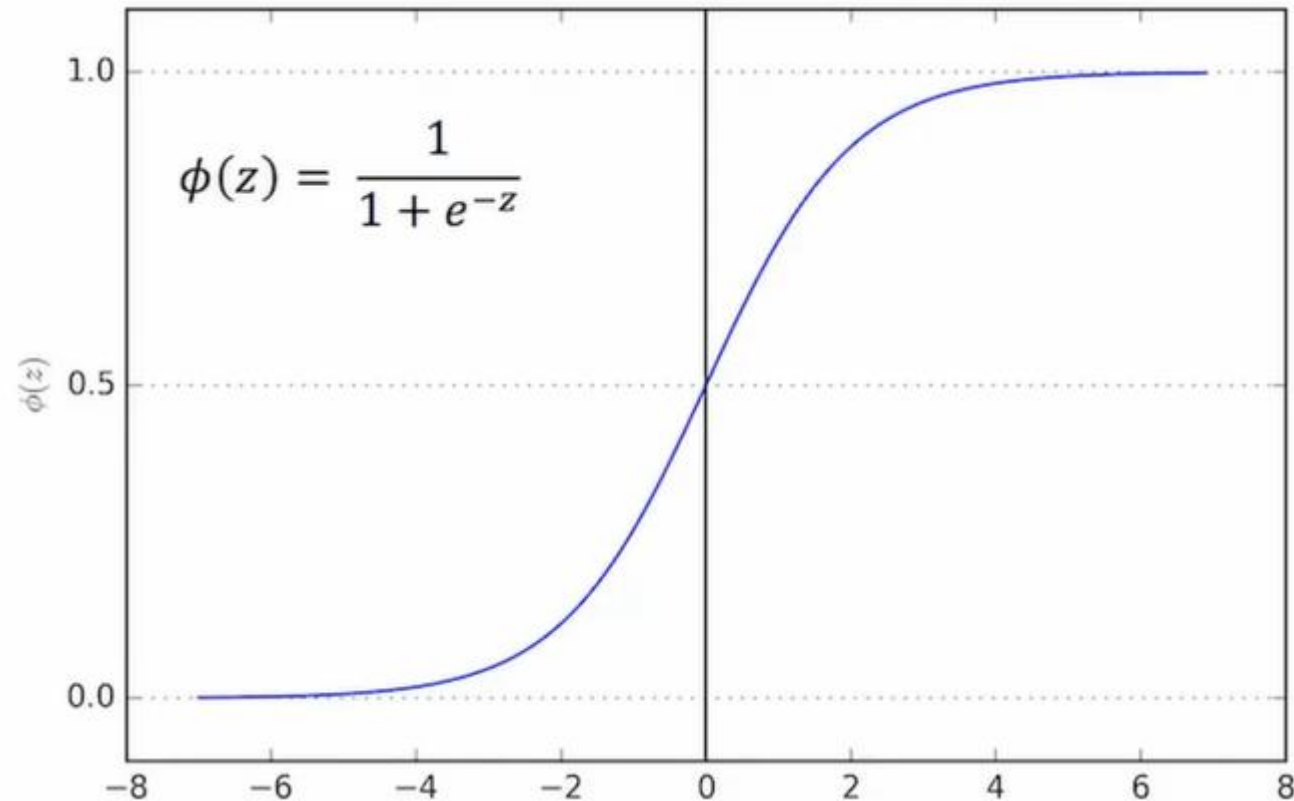
Classification

Instead we can transform our linear regression to a logistic regression curve.



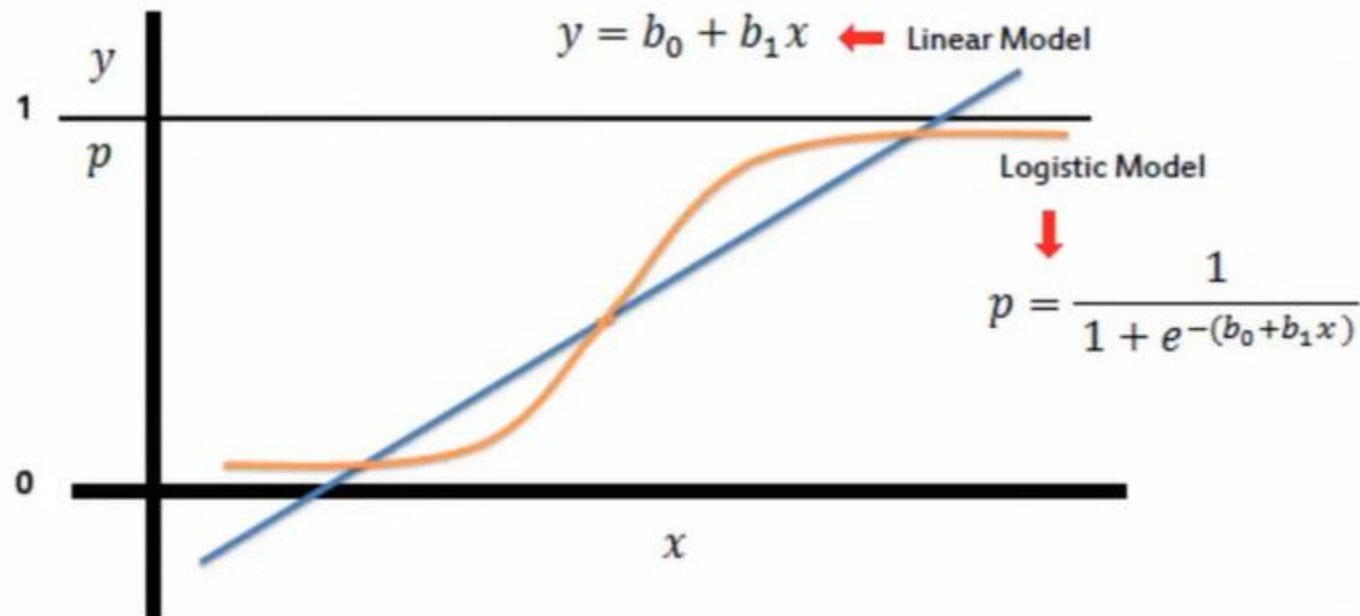
Classification

The Sigmoid (aka Logistic) Function takes in any value and outputs it to be between 0 and 1.



Classification

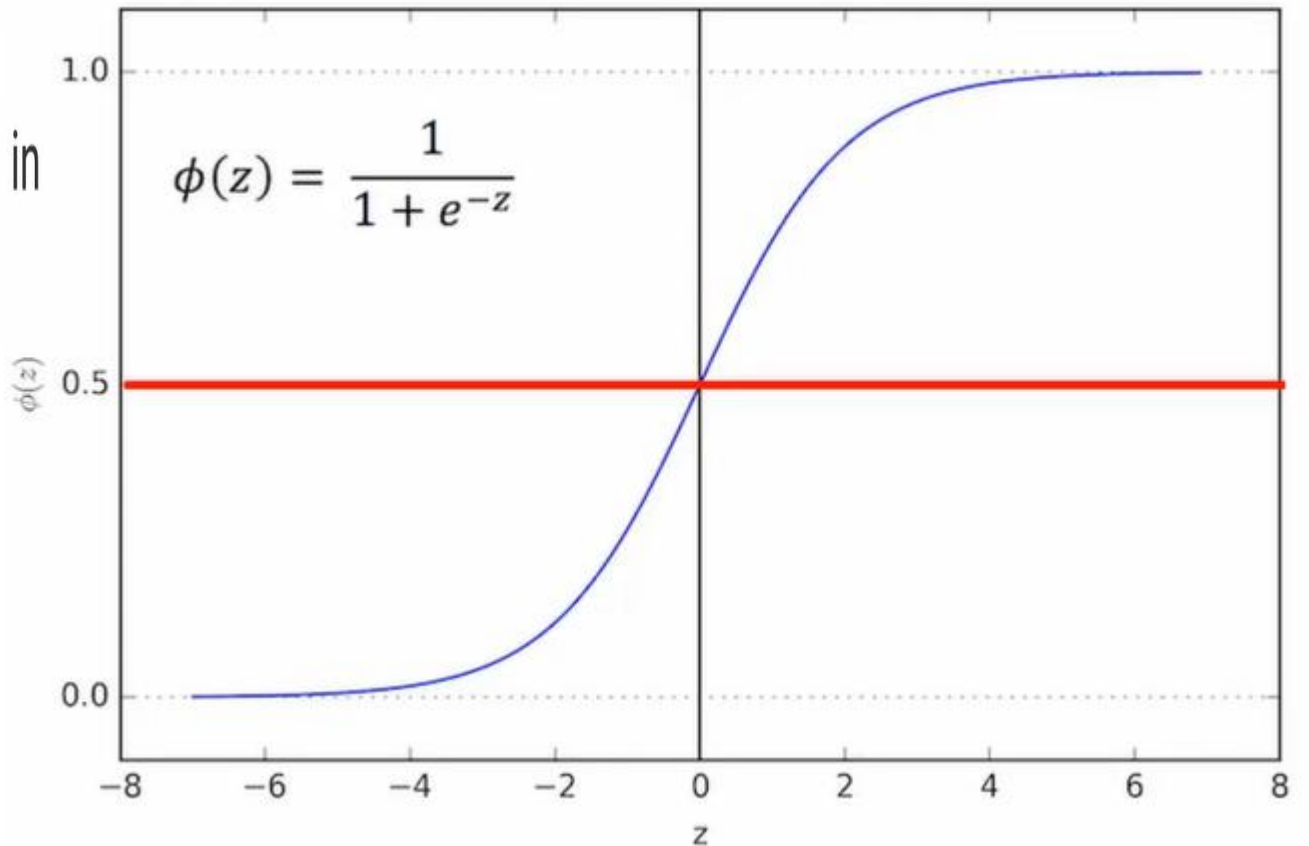
This means we can take our Linear Regression Solution and place it into the Sigmoid Function.



Classification

This results in a probability from 0 to 1 of belonging in the 1 class.

We can set a cutoff point at 0.5, anything below it results in class 0, anything above is class 1.



Classification

We can use a confusion matrix to evaluate our model.
For example, imagine testing for disease.

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

Example: Test for presence of disease
NO = negative test = False = 0
YES = positive test = True = 1

Classification

n=165		Predicted: NO	Predicted: YES	
Actual: NO		TN = 50	FP = 10	60
Actual: YES		FN = 5	TP = 100	105
		55	110	

Basic Terminology:

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

Classification

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Accuracy:

- Overall, how often is it **correct**?
- $(TP + TN) / \text{total} = 150/165 = 0.91$

Classification

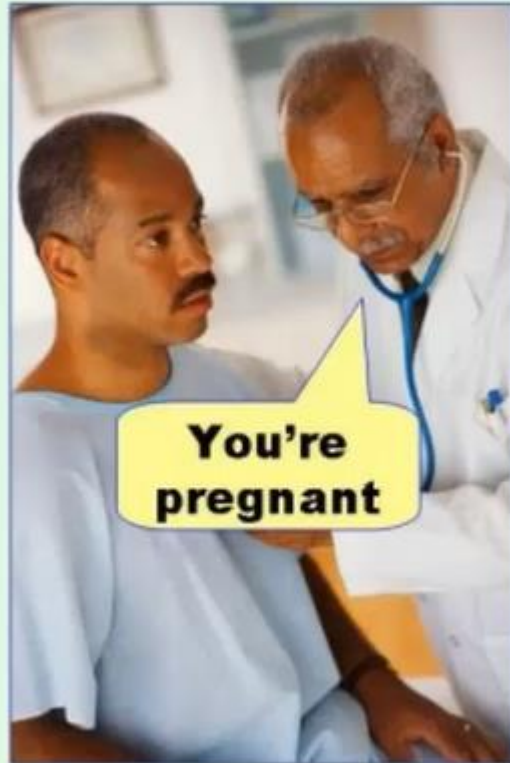
n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Misclassification Rate
(Error Rate):

- Overall, how often is it **wrong**?
- $(FP + FN) / \text{total} = 15/165 = 0.09$

Classification

Type I error
(false positive)



Type II error
(false negative)

