Session: Logistic Regression

Logistic Regression
Decision Boundary
Advanced

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Problem: Applying for Credit Card

- In the US, most adults have a credit score (a.k.a. FICO score)
- Ranges from 300 to 850
- Credit score important for loans, mortgages, and credit cards

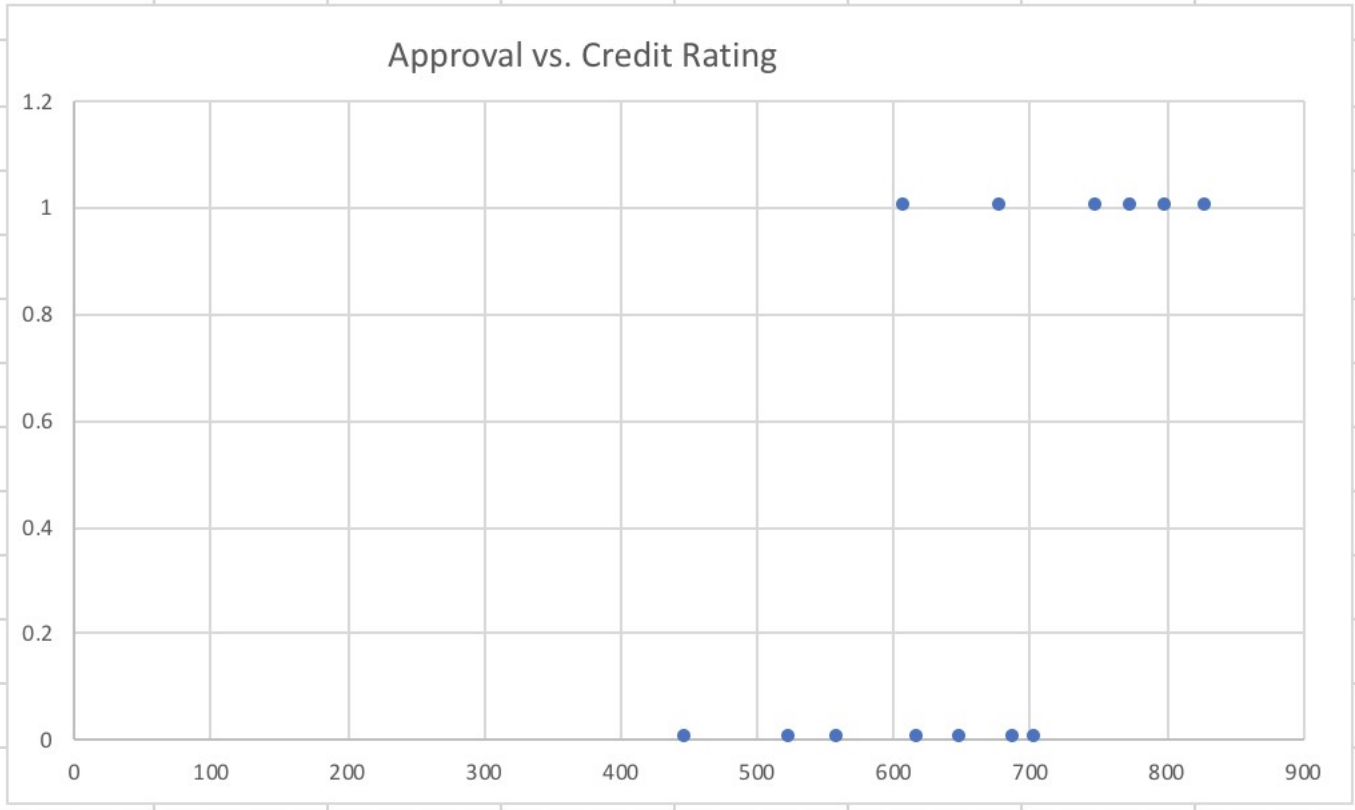
Problem: Applying for Credit Card

- Data:
  - X – Credit Score
  - Y – Approval (Yes/No)
- Can we predict approval?

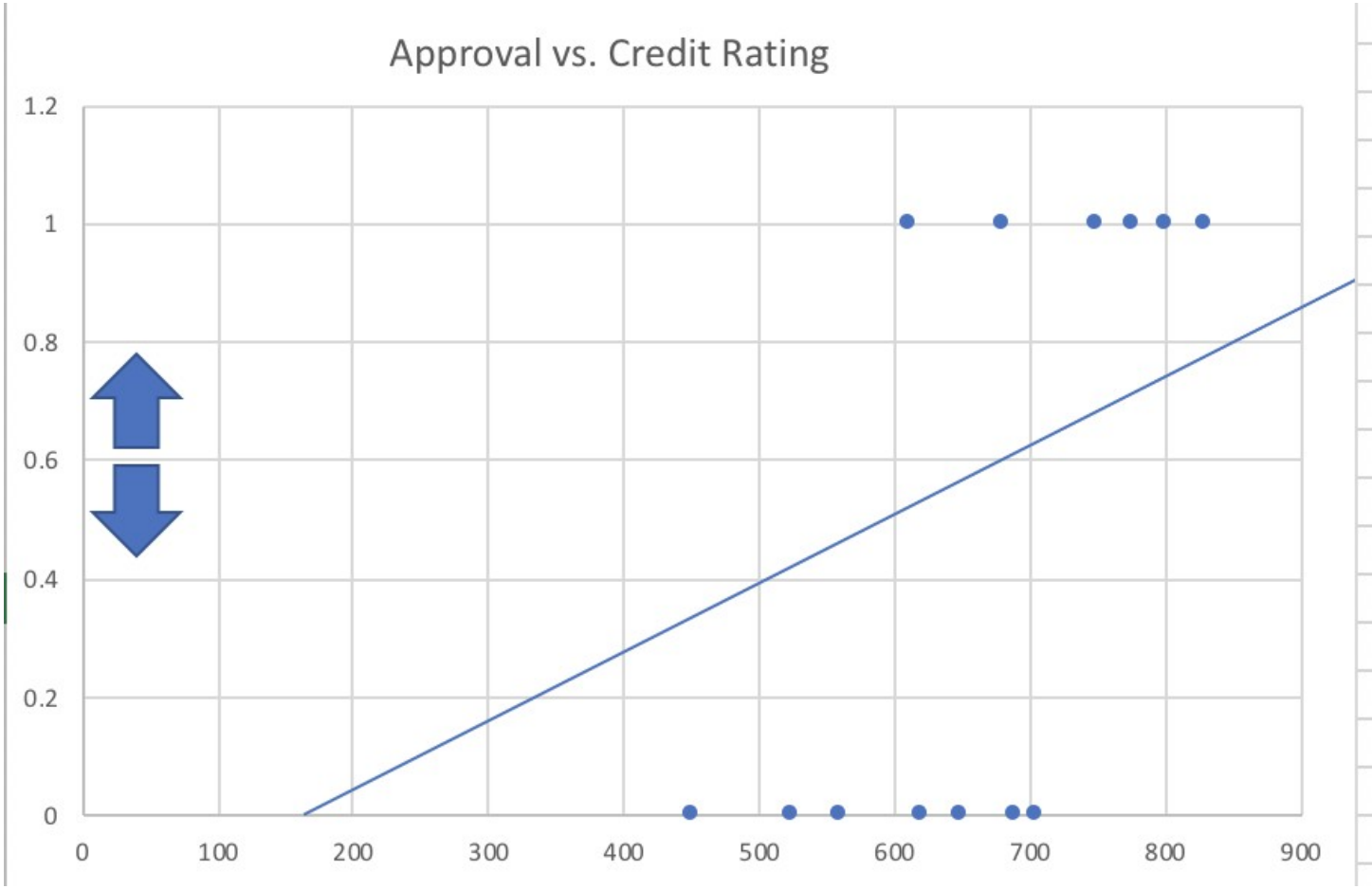
Credit Score	Approved?
560	No
750	Yes
680	Yes
650	No
450	No
800	Yes
775	Yes
525	No
620	No
705	No
830	Yes
610	Yes
690	No

Visualize the Credit Approval Data

	A	B
1	560	0
2	750	1
3	680	1
4	650	0
5	450	0
6	800	1
7	775	1
8	525	0
9	620	0
10	705	0
11	830	1
12	610	1
13	690	0

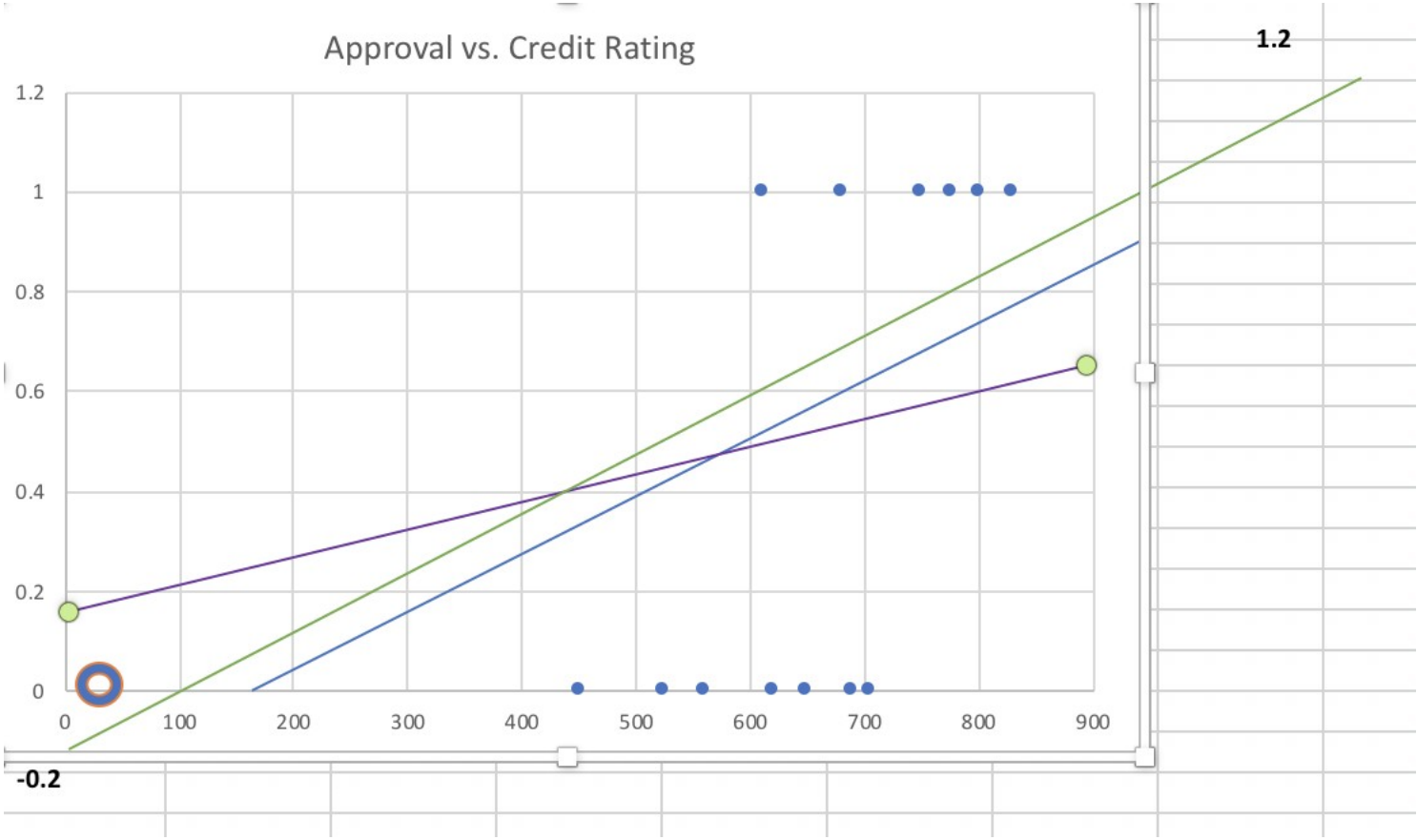


Try Linear Regression



- But…

Problems with Applying Linear Regression



- Unstable
- Wrong value range

Logistic Regression

➔ **Logistic Regression**

Decision Boundary

Advanced

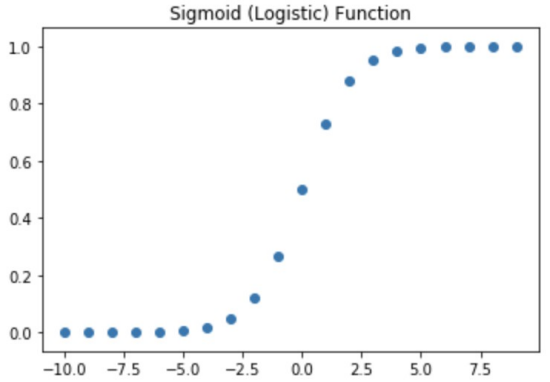
Sigmoid (Logistic) Function

- Instead of

$$H = T_0 + T_1 x_1 + T_2 x_2 + \ ... + T_n x_n$$

- Let us change our hypothesis to
- Let us change our hypothesis to

$$S(H) = \frac{1}{1 + e^{-H}}$$

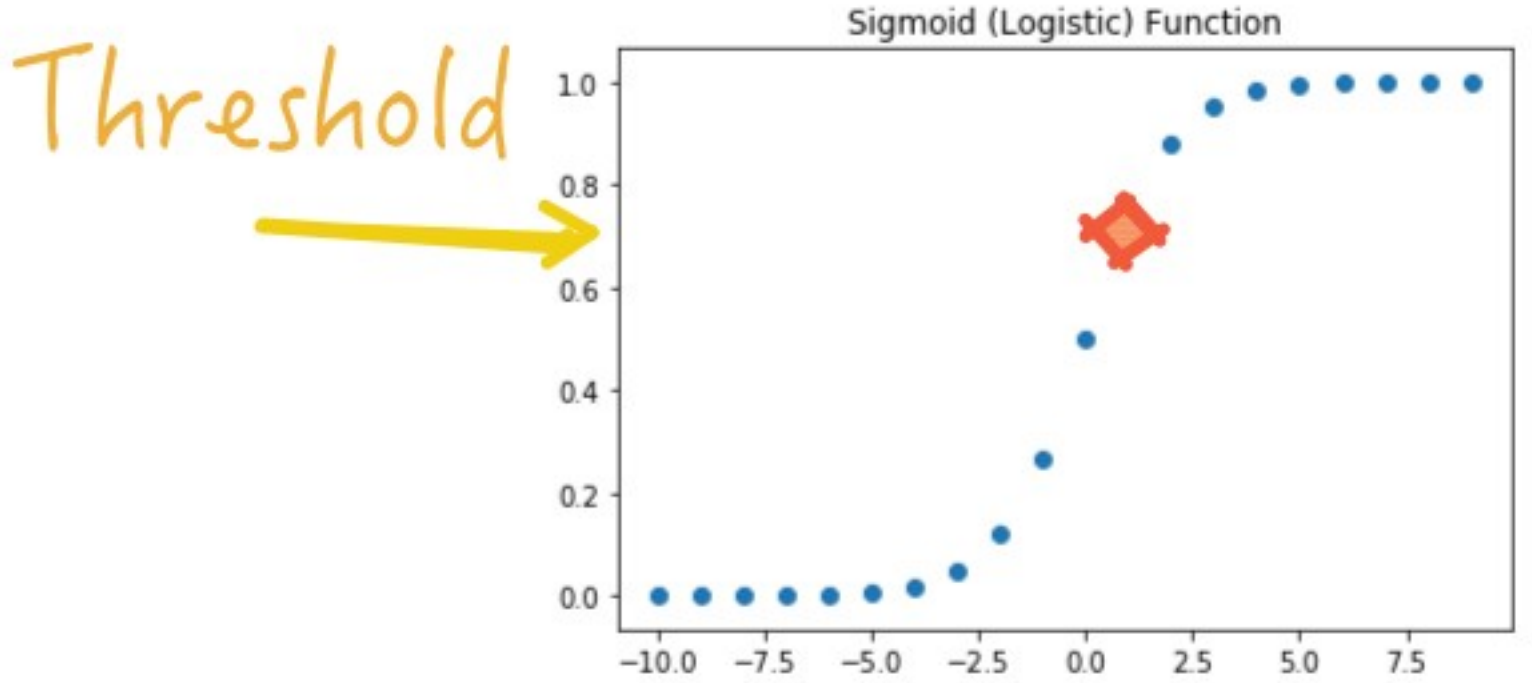


- So, instead of straight line, we get an S-shape

- How do we get the best values of [T_0, T_1, ..., T_n]?
  - = It is also Gradient Descent
  - = But first, let us describe how we use the result

Answer Provided by Logistic Regression

- Probability p that the answer is 1
- p > 0, p < 1
- p > 0.5 => answer is **yes**
- p < 0.5 => answer is **no**



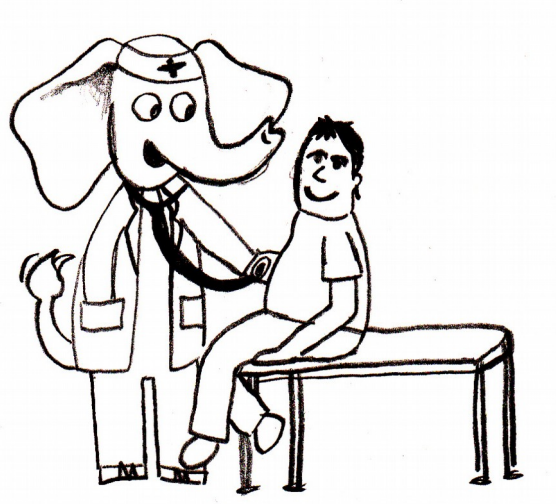
Logistic Regression Use Cases

- In ML Spark, this is simply
  - `LogisticRegression.setThreshold`

- But in life it may be
  - Fraud investigation
  - Tumor prediction – malignant or benign
  - Credit approval
  - Student admission
  - Character recognition

How to Explain Predictions

- Say, you get 75% for malignant tumor prediction
  - Tell the patient that the probability of him having cancer is 75%
- But what should the doctor do?
- ”Conservative” doctor
  - Operates at 25%
- “Progressive” doctor
  - Recommends “wait and see”
- And both can explain why they are right



Lab: Logistic Regression

- **Overview**:
  Practice Logistic Regression
- **Approximate Time**:
  30 mins
- **Instructions**:
  Follow appropriate Python / R / Spark instructions
  - **LOGIT-1: Credit card approval**

Decision Boundary

Logistic Regression

➔ **Decision Boundary**

Advanced

The Meaning of Sigmoid

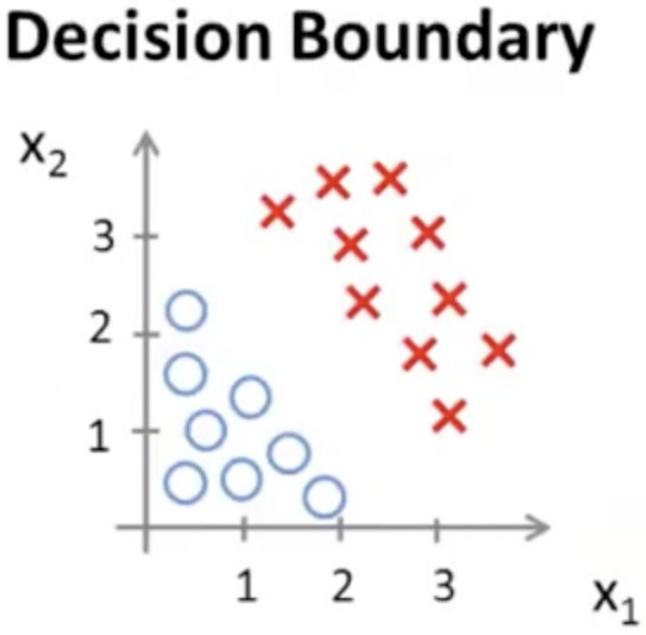
- Sigmoid

$$S(z) = \frac{1}{1 + e^{-z}}$$

- Z = 0 => S(z) = ½
- Z > 0 => S(z) > ½
- Z < 0 => S(z) < ½

Decision Boundary

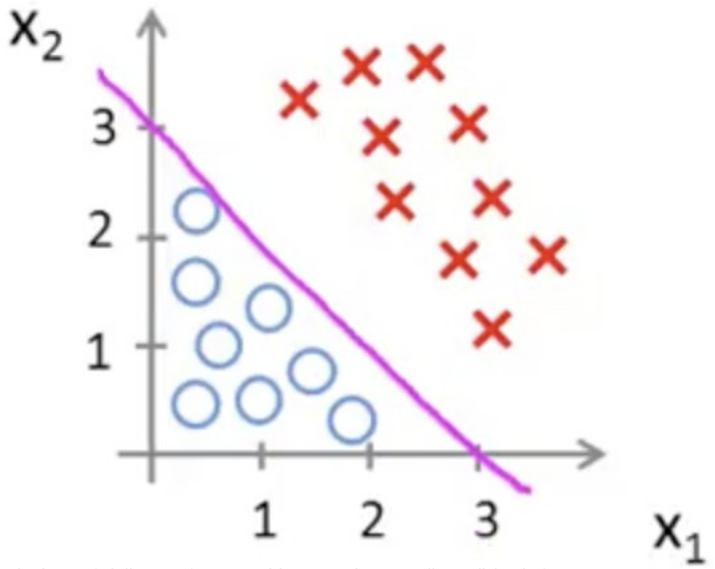
- Divides the “yes” answer from the “no” answer



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

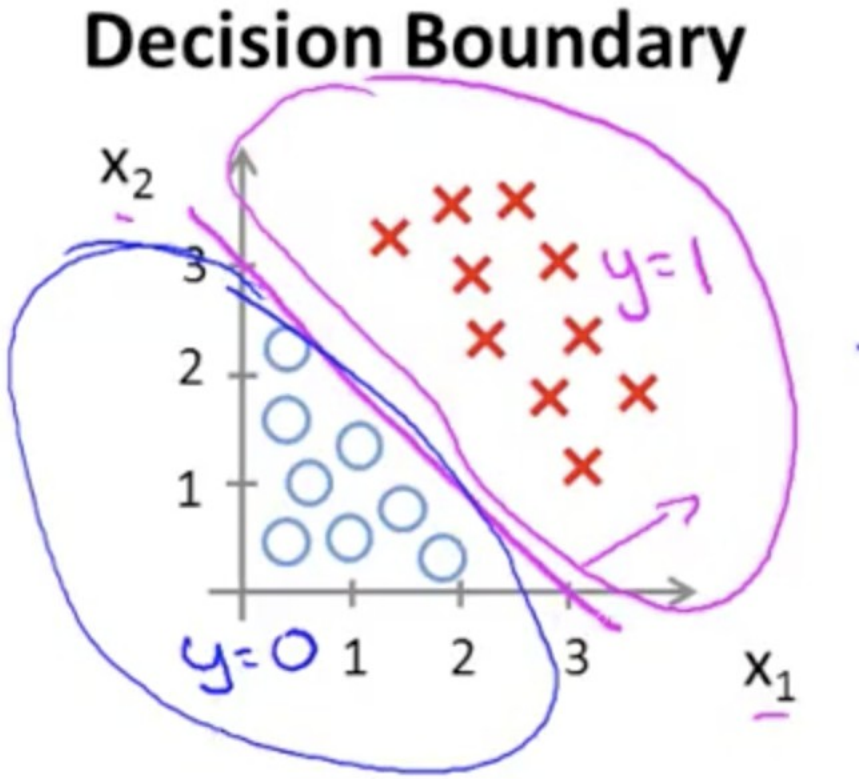
Decision Boundary in Numbers

- Say
  - T0 = -3
  - T1 = 1
  - T2 = 1
- We will predict
  - Y = 1 if
  - -3 + X1 + X2 >= 0



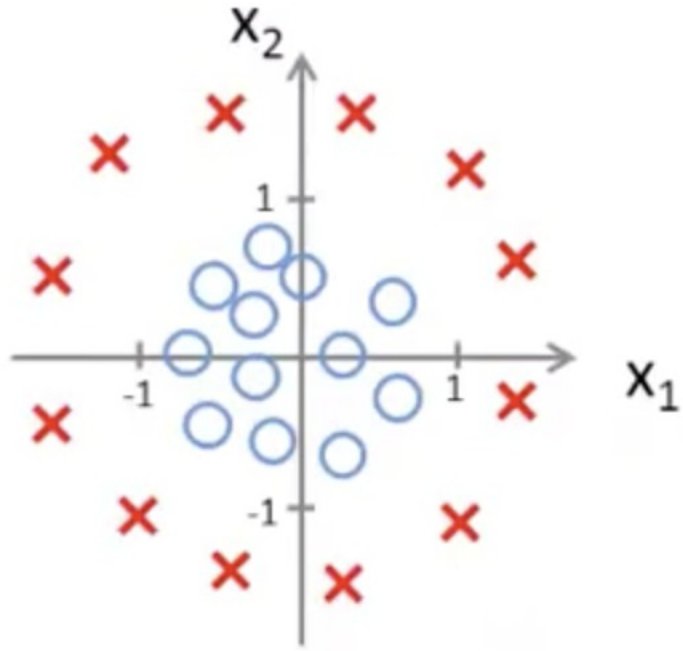
Decision Boundary - Regions

- Decision boundary separates our points into 2 groups



Non-linear Decision Boundary

- Sometimes, points cannot be cut up by a line



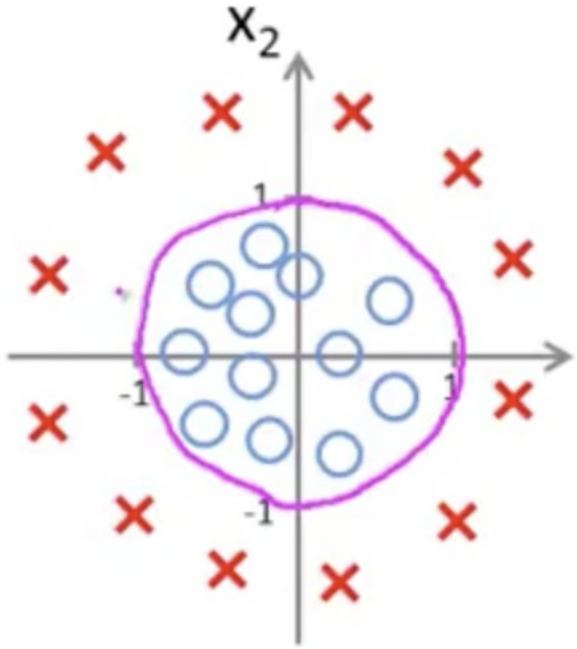
Adding New Features

X1	X2	X3 = X1**2	X4 = X2**2
2	3	4	9
1	4	1	16
3	5	9	25

- My formula
- T(X) = T0 + T1*X1 + T2*X2 + T3*X3 + T4*X4
- Let us choose T0 = -1; T1 = T2 = 0
- Predict y = 1 if -1 + x1**2 + X2**2 > 0

Equation of a Circle

- New formula
- -1 + x1**2 + X2**2 > 0 or
- X1**2 + X2**2 > 1



What Shapes are Possible?

- Circles
- Ellipses
- Many other shapes with >2 degree polynomials

Advanced

Logistic Regression

Decision Boundary

➔ **Advanced**

Multinomial logistic regression

- We have seen Logistic Regression predicting binary outcomes
  - Approved / Denied
- We can use it to calculate 'more than two' states as well
  - multinomial logistic regression

- For K possible outcomes
  - Chose one outcome as a “pivot”
  - The other K−1 outcomes can be separately regressed
    - against the pivot outcome

Solving the Logistic Regression Problem

- Earlier we said that training the Logistic Regression model is done with Gradient Descent
- But what is our Cost Function?

What We Used for Linear Regression

- Cost function for Linear Regression
- Cost function for Linear Regression

$$C(T_0, T_1, , , Tn) = \frac{1}{2m} \sum_{1}^{m} (y'_i - yi)$$

- But then we added a sigmoid to our prediction
- So now the cost function is not convex
- But then we added a sigmoid to our prediction
- So now the cost function is not convex

Cost Function for Logistic Regression

- To make it convex
- We will use a different cost function

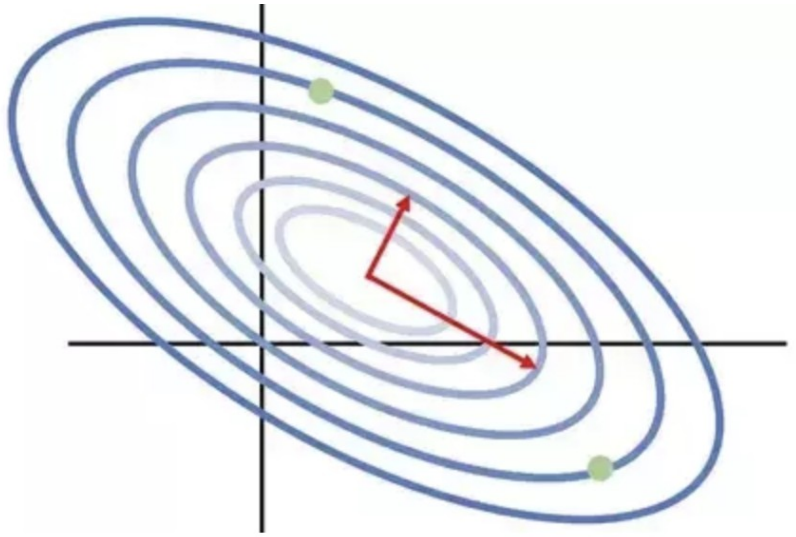
$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$\text{Cost}(h_\theta(x), y) = -\log(h_\theta(x)) \qquad \text{if y} = 1$$

$$\text{Cost}(h_\theta(x), y) = -\log(1 - h_\theta(x)) \qquad \text{if y} = 0$$

Advanced Optimization

- Conjugate gradient
  - For symmetric, positive mx



- BFGS
  - Authors' names
- L-BFGS
  - Modified
  - For limited-memory, sparse matrices

Evaluating Classification Models

- Let's consider a binary classifier
  - Picks one of two outcomes (spam / not-spam)
- Two approaches
  - Confusion matrix
  - ROC curve

Confusion Matrix / Error Matrix

- Let's consider a binary classifier
  - Picks one of two outcomes (spam / not-spam)
- Say we are classifying 10 emails (6 spam, 4 not-spam)
- See '**confusion matrix**' below

Classification =>	Spam	Not Spam
Actual Spam (6 total)	4 count 66% accuracy Correct **True-Positive Rate (sensitivity)**	2 count 33% Incorrect **False negative rate (miss rate)**
Actual Not Spam (4 total)	1 count 25 % Incorrect **False-positive rate (fall-out)**	3 count 75% Correct **True negative rate (specificity)**

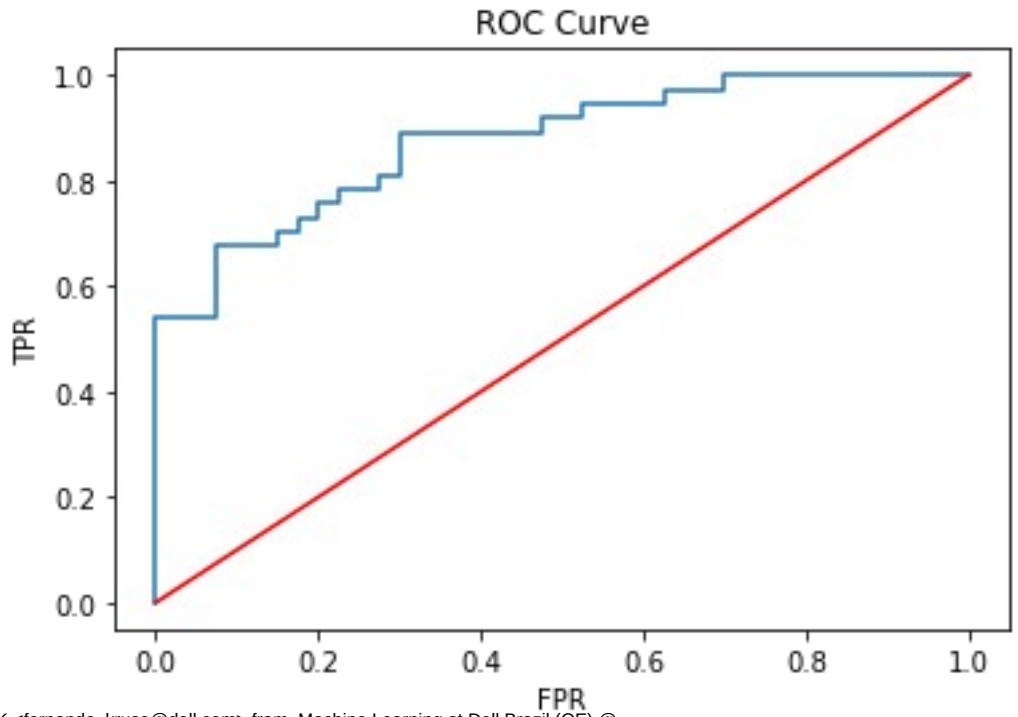
Measuring Accuracy of Logistic Model

- Since Logistic Regression is used for classification we can use
  - Confusion Matrix
  - ROC and AUC (Area Under Curve)
- Confusion Matrix:
  correct: 14 + 5 = 19
  missed: 3 + 1 = 4
  accuracy = 19/(19+4) = 82.6%

```
+-----+---+---+
|admit|0.0|1.0|
+-----+---+---+
|    0| 14|  3|
|    1|  1|  5|
+-----+---+---+
```

ROC Curve

- ROC – diagnostic capability of a binary classifier
  - With changing threshold
- True Positive Rate (TPR) vs False Positive Rate (FPR)
- AUC = 0.874



Lab: Logistic Regression

- **Overview**:
  Practice Logistic Regression
- **Approximate Time**:
  30 mins
- **Instructions**:
  Follow appropriate Python / R / Spark instructions
  - **LOGIT-2: College Admission**

Review Questions

- How does Logistic Regression differ from Linear Regression?
- What do they have in common?

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12