Logistic Regression

Week 05 - Day 01

Linear Regression

Lasso Regression

Ridge Regression

Elastic Net

Linear Regression - regression

Lasso Regression - regression

Ridge Regression - regression

Polynomial Regression - regression

Logistic Regression - ???

Linear Regression - regression

Lasso Regression - regression

Ridge Regression - regression

Polynomial Regression - regression

Logistic Regression - **classification** 

Logistic Regression = binary classification

Binary Classification with 1 feature

tinder

Sean Rad, 29

Verified

Like Me On Tinder

tinder

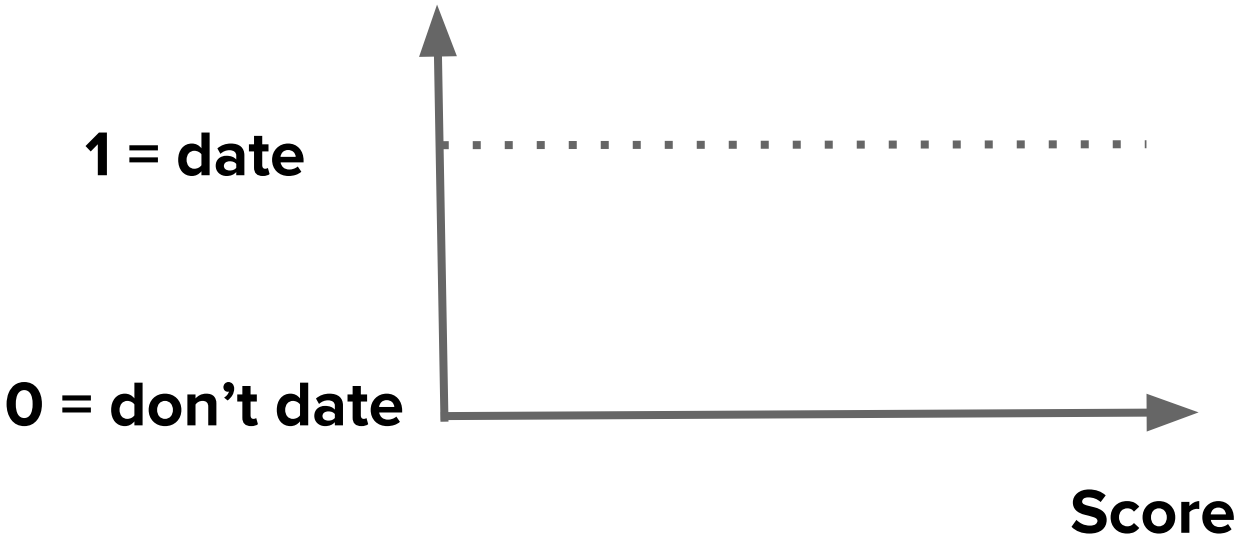
Rosette, 32

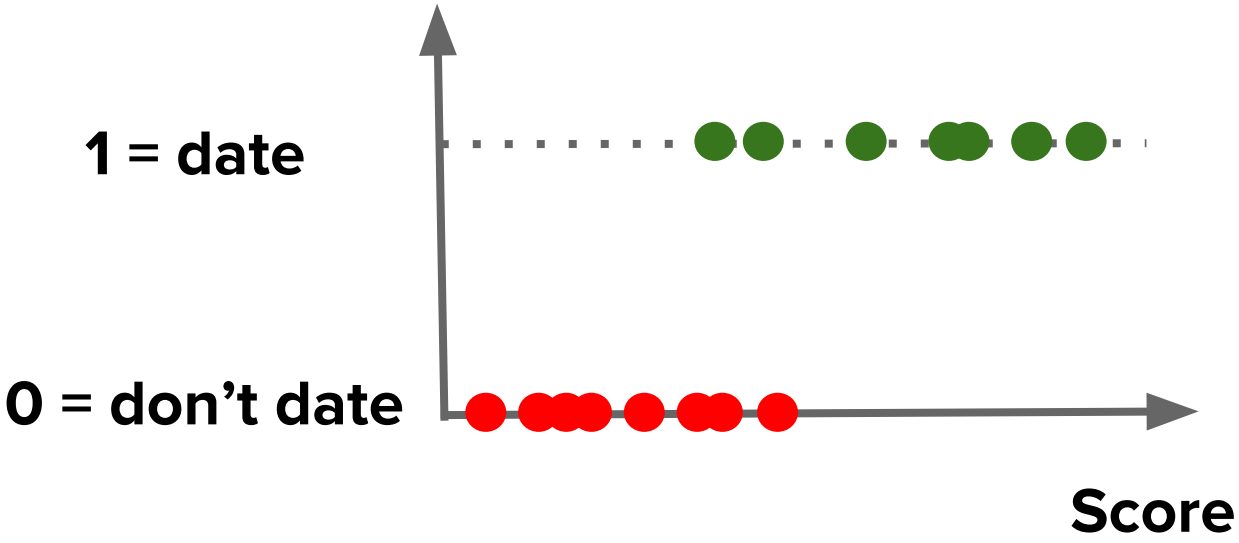
Vice President, Global Communications & Bra...

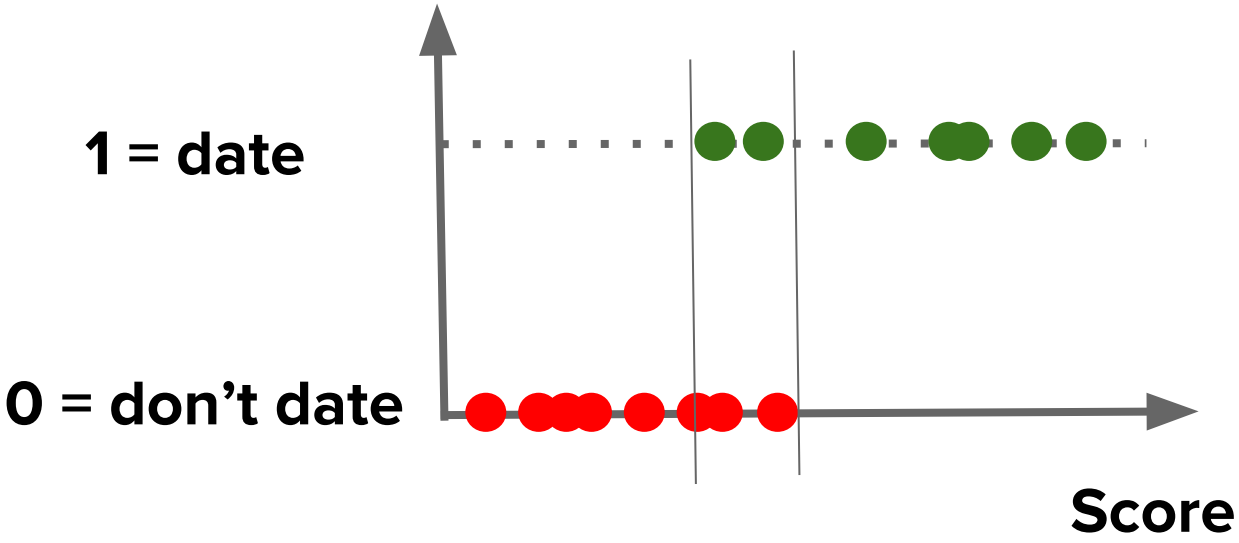
Like Me On Tinder

score = hotness + 2*personality

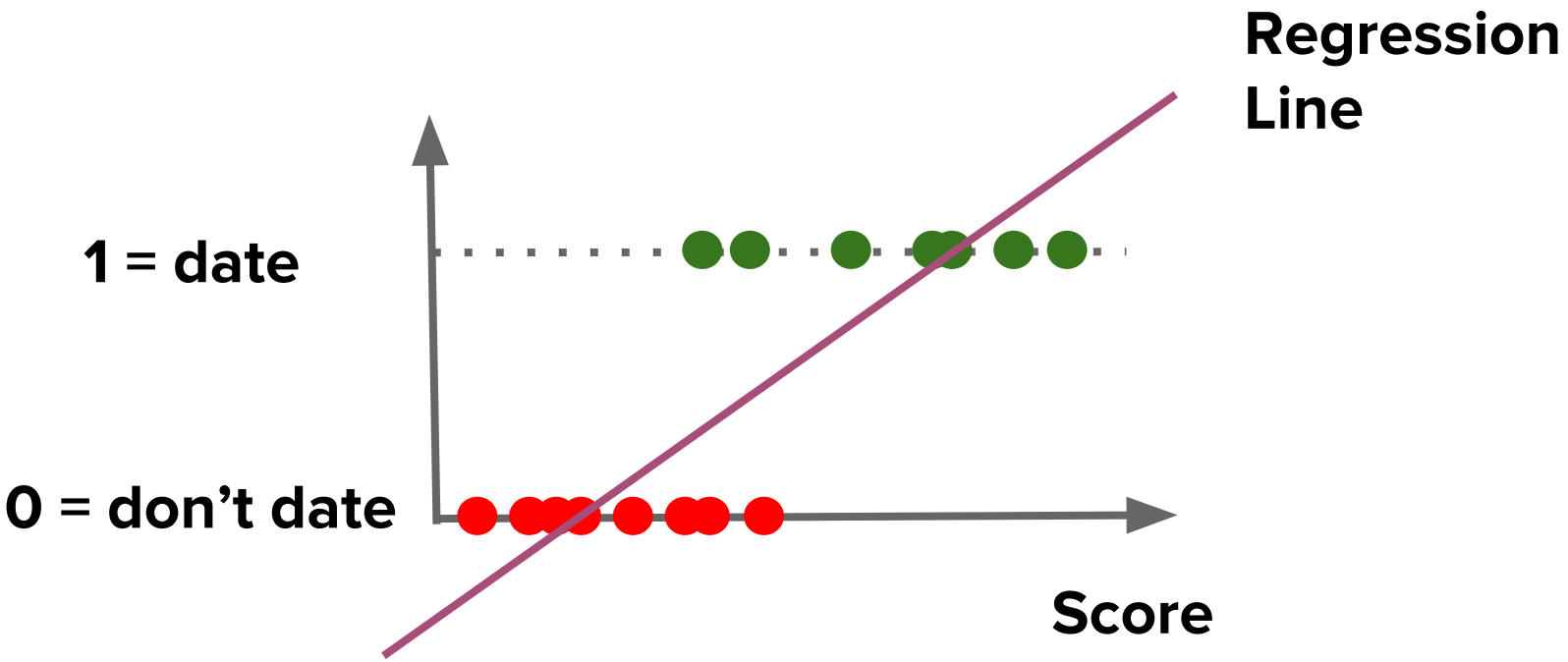
#	Score	Date
1	8.2	yes
2	5.8	no
3	6.2	no
4	6.1	yes


1 = date
0 = don’t date
Score


1 = date
0 = don’t date
Score


1 = date
0 = don’t date
Score

Linear Regression


Regression Line
1 = date
0 = don’t date
Score

Problems?

Problem: values >1 or <0

Solution: ???

Problem: values >1 or <0

Solution: bound y to [0,1]

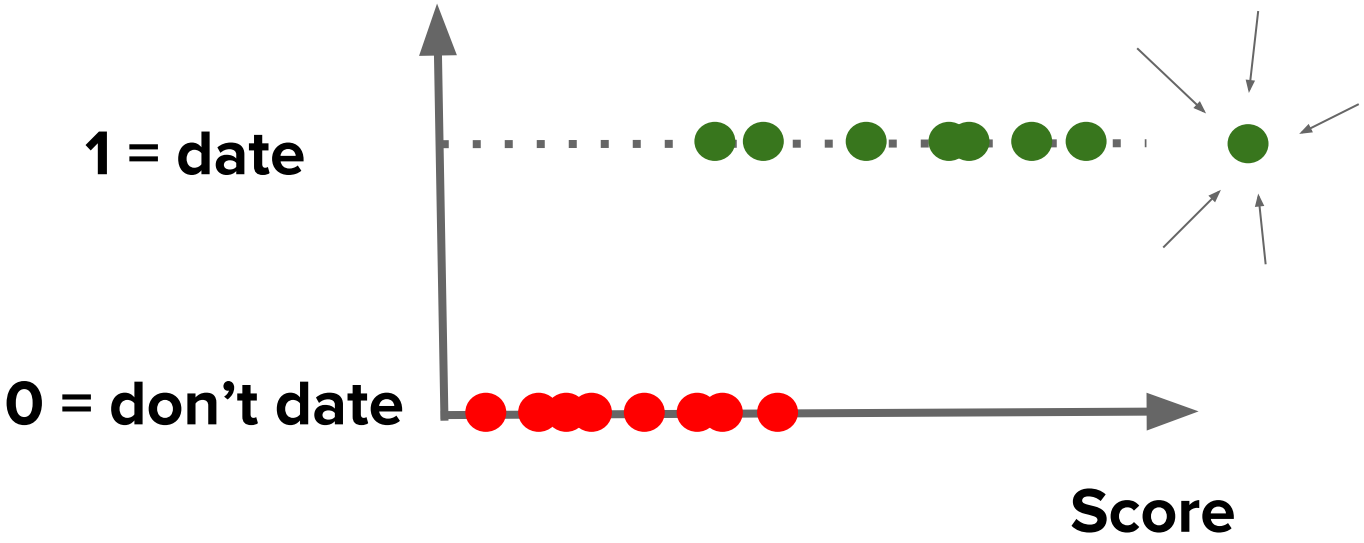
Problem: values between 0 and 1

Solution: ???

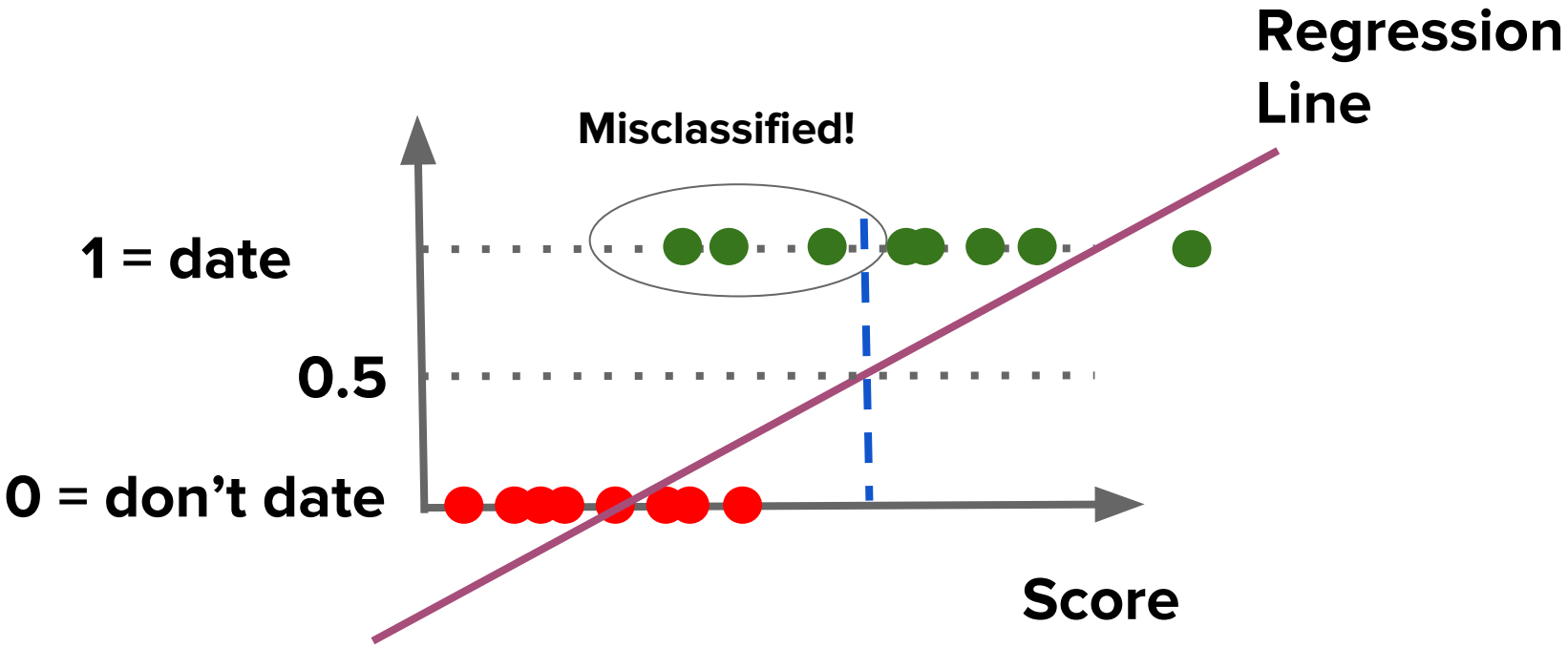
Problem: values between 0 and 1

Solution: use 0.5 as cutoff value

A new point


1 = date
0 = don’t date
Score

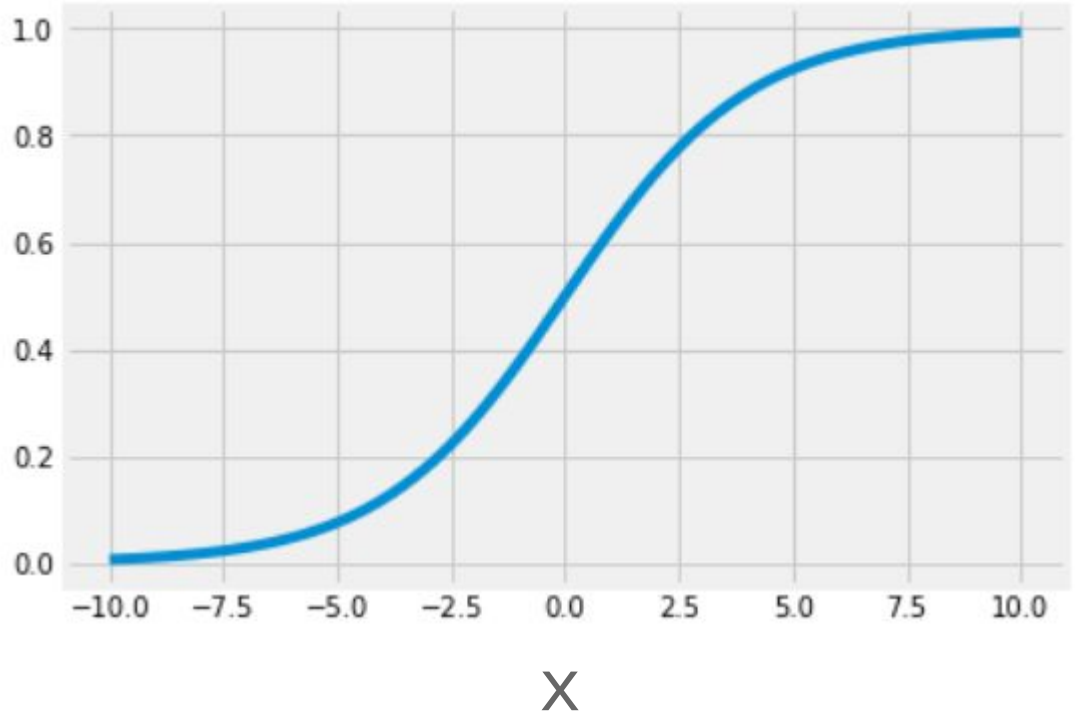
New regression line?


Regression Line
Misclassified!
1 = date
0.5
0 = don’t date
Score

Solution: logistic function

Logistic Function

$1 / (1 + e^{**}(-x))$



Notebook Plot_logit_function.ipynb

Coefficients Interpretation

$$\Pr(Y = 1) = \frac{1}{1 + \exp(-[-3.92 + 0.014 \times (\text{gender})])}$$

exp(a) = e**a

https://stats.stackexchange.com/questions/20996/how-to-interpret-coefficients-from-a-logistic-regression

What's the interpretation of this coefficient?

$$\Pr(Y = 1) = \frac{1}{1 + \exp(-[-3.92 + 0.014 \times (\text{gender})])}$$

*With "gender", if Female = 0 and Male = 1, you can assert that the odds of your outcome for women are e**(0.014) = 1.01 times that of the odds of your outcome in men*

Odds Ratios

P = probability

Odds-ratio(p) = p/(1-p)

p	1-p	Odds ratio p/(1-p)
0.1	0.9	0.11
0.25	0.75	0.33
0.5	0.5	1
0.75	0.25	3
0.9	0.1	9

https://www.youtube.com/watch?v=5zPSD_e_N04

Find the best beta1 and beta0

No simple formula with derivation

=

Optimization process

Strange loss function

(logistic loss)

From p to class

Standard cutoff = 0.5

Another parameter to tune!

Why is it called Regression?

We’re just predicting a number

(between 0.0 and 1.0)

Summary

- Logistic regression = binary classification

- Logistic regression = binary classification
- Map all the points in the interval [0,1] with logistic function

- Logistic regression = binary classification
- Map all the points in the interval [0,1] with logistic function
- Use cutoff, usually 0.5

- Logistic regression = binary classification
- Map all the points in the interval [0,1] with logistic function
- Use cutoff, usually 0.5
- The cutoff can be tuned!

- Logistic regression = binary classification
- Map all the points in the interval [0,1] with logistic function
- Use cutoff, usually 0.5
- The cutoff can be tuned!
- Interpretation of coefficients = odds ratio