1

1

Logistic Regression

2

- The statisticians approached the problem - "how can we use linear regression to solve this?"
  - We could consider the following encoding
    - Dependent variable coded as 0 (Not Spam) or 1 (Spam)

2

Logistic Regression

3

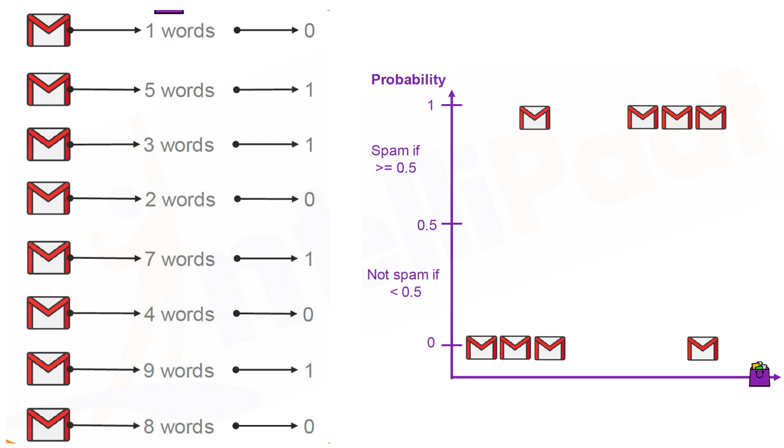
- We can fit a linear regression to this binary response
  - and classify as Spam if $\hat{y}$ > 0.5 and Not Spam otherwise, interpreting $\hat{y}$ as a probability that Email is Spam

3

Logistic Regression

4

- Spam email classifier



4

Logistic Regression

5

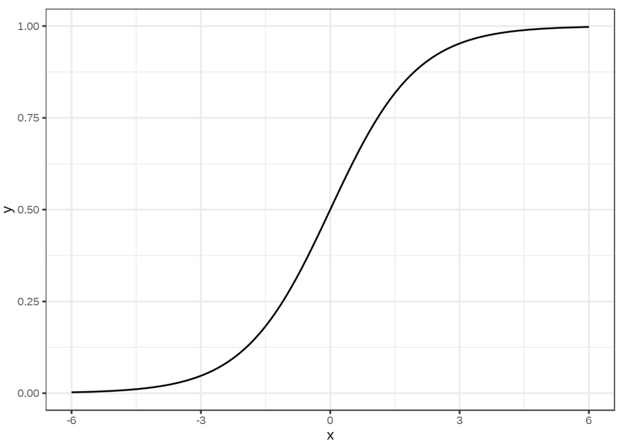
- A major problem with such an approach
- Linear regression models produce values in (-∞, +∞), which does not make sense as a probability
  - Employ a function that constrains the values between 0 and 1

5

Logistic Regression

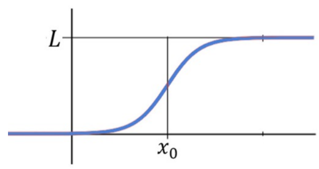
6

- Logistic function (Sigmoid)



$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

x_0 = x value of midpoint
L = maximum value
k = growth rate



6

Logistic Regression

7

- Logistic regression model

Sigmoid input

Linear regression output

$$\log\left(\frac{p_X}{1-p_X}\right) = \beta_0 + \beta_1 X$$

$$p_X = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \quad 0 \le p_x \le 1$$

7

Logistic Regression

8

- Terminologv
  - The logit $\log\left(\frac{p_X}{1-p_X}\right)$ Log odds
  - The odds of an event is defined as

$$\text{odds}(Y=1) = \frac{P(Y=1)}{1-P(Y=1)} = \frac{p}{1-p}$$

- Chance can be expressed either as a probability or as odds

8

Logistic Regression

9

- Ranges

Measure	Min	Max	Name
$P(Y=1)$	0	1	"probability"
$\frac{P(Y=1)}{1-P(Y=1)}$	0	∞	"odds"
$\log\left[\frac{P(Y=1)}{1-P(Y=1)}\right]$	$-\infty$	∞	"log-odds" or "logit"

9

Logistic Regression

10

- The cost/loss function (compares the model's predictions to the target values)
  - The likelihood function

$$\prod_{i=1}^{N} p_{x_i}^{y_i}(1-p_{x_i})^{1-y_i}$$

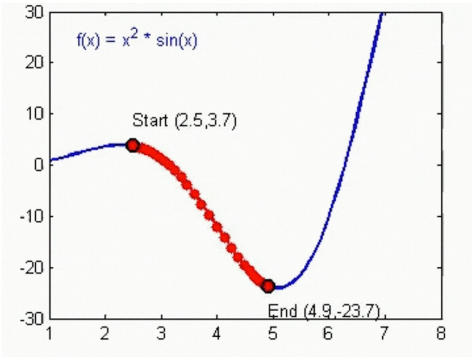
  - The likelihood is a function of model parameters, and we can estimate them by maximizing the likelihood
    - Maximum likelihood estimates (MLE)

10

Logistic Regression

11

- The cost/loss function
  - No closed form solution for MLE
  - We rely on numerical approximation to find the MLE
    - Most software uses the gradient descent algorithm



11

Logistic Regression

12

- In practice, it is more convenient to maximize the log of the likelihood function
  - product of a large number of small probabilities can easily lead to underflow in computing machines

12

Thank You!

13