Machine Learning Fundamentals

Practical Machine Learning (with R)

UC Berkeley

Fall 2015

SESSION 3 – LOGISTIC REGRESSION

Topics

- Administrativa
- Quiz
- Review and Q&A
  - *Using git / github*
- New Topics

ADMINISTRATA

Assignments

- Google Group:
  - https://groups.google.com/forum/#!forum/**csx460**
- Github Group:
  - https://github.com/**csx460**
- Create a Github Account
- Clone assignments from Github
- Commit and Push Answers
- (Send Pull Request)
- → Commit assignments to github now

REVIEW

Expectations

- Understand 3 Things That All ML Algorithms Share
- Understand the difference between the ML algorithm and the function that the ML algorithm produces
- Use git as version control
- Understand how to use `lm` for creating linear regression problems

Git / GITHUB / Source Tree

- **Workflow**
  - clone
  - branch
  - (work)
  - add
  - commit (early and often)
    - tag
  - push
  - Also checkout, status, log

3 Requirement for Algorithm

- A method for evaluating how well the algorithm performs (**ERRORS**)
- A restricted class of function (**MODEL**)
- A process for proceeding through the restricted class of functions to identify the functions (**SEARCH/OPTIMIZATION**)

Review

- https://github.com/CSX460

- git / github / Rstudio + git

QUIZ

Linear Regression (Significance)

`Pr(>|t|)`

Linear regression t-statistic is the probability that the "true value" of the statistic falls outside the student t-distribution.

- Is expressed as a probability.
- Lower is "better" i.e. more significant

Think of it (loosely) as the probability of the coefficient being off.

LINEAR REGRESSION (INTUITION)

- Coefficients … multiply then sum

- Number Line (in units of the response)
  - Start at intercept
  - Multiple term by value of the variable
  - Move those number of units left or right.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   51.3541     0.4593 111.814  < 2e-16 ***
EngDispl      -3.7454     0.2507 -14.941  < 2e-16 ***
NumCyl        -0.5880     0.1722  -3.414 0.000664 ***
```

Linear Regression Errors

- Two different types of errors measured
  - For ***fitting*** models
  - For **comparing** models

- Minimize square error loss (SSE) ***sum of squared errrors***

$$argmin_{\boldsymbol{\beta}}\left(\sum(\hat{y}-y)^2\right)$$

- *choose Beta such that the sum of squared errors is minimized.*
- Assumes there is a solution

Transformations

- Centering and Scaling: `scale`*
- Resolve skewness: `log, sqrt, inv`
- Resolve outliers: spatial sign, `PCA`

Some algorithms require scaling

Some are insensitive

Time consuming

Somewhat of an art

- Genetic algorithms (GA)

NEW TOPICS

Classification of Machine Learning Algorithms

- Errors
- Restricted Class of Functions
- Search Methodology

CLASSIFICATION OF MACHINE LEARNING ALGORIHMS

- Data: Output / Response
  - Regression (continuous) vs
  - Classification (categorical)

  Presense of Response:
  - SUPERVISED LEARNING
    Known previous outcomes: "Labelled"
  - UNSUPERVISED LEARNING
  - Unknown previous outcome: "Unlabelled"

  - Adaptive Learning
    - Reinforcement

Classification of Machine Learning Algorihms

- Data: Inputs
  - Type excepted
  - How handled

Example of ML algorithm(s)

- Spam Filter
- handwriting recognition (svm)
- Traffic engineering (lights)
- Weather prediction
- Sentiment analysis (social media)
- Netflix Recommender
- Fraud detection (Visa)
- Imaging processing
- (network) Intrution detection
- Self-driving cars

LOGISTIC REGRESSION

Background

Categorical Modeling:

$$\hat{y}_{cat} = f(\vec{x})$$

- Inputs
  - Categorical
  - Continuous variable can assume any value

Outputs:

How do we handle categories?

- same as linear regression?

Background

- Errors!

$$\hat{\boldsymbol{y}}^{\boldsymbol{cat}} \neq \boldsymbol{y}$$

  - Problem …

$$argmin_{\boldsymbol{\beta}} \sum \begin{Bmatrix} 1 \mid \hat{y} \neq y \\ 0 \mid \hat{y} = y \end{Bmatrix}$$

FUNCTION …

- Do the easiest thing first …
  Start with 2 categories “binomial dist”
  - A | B
  - TRUE | FALSE
  - 0 | 1

“Looks Math-y”

Need a tool …

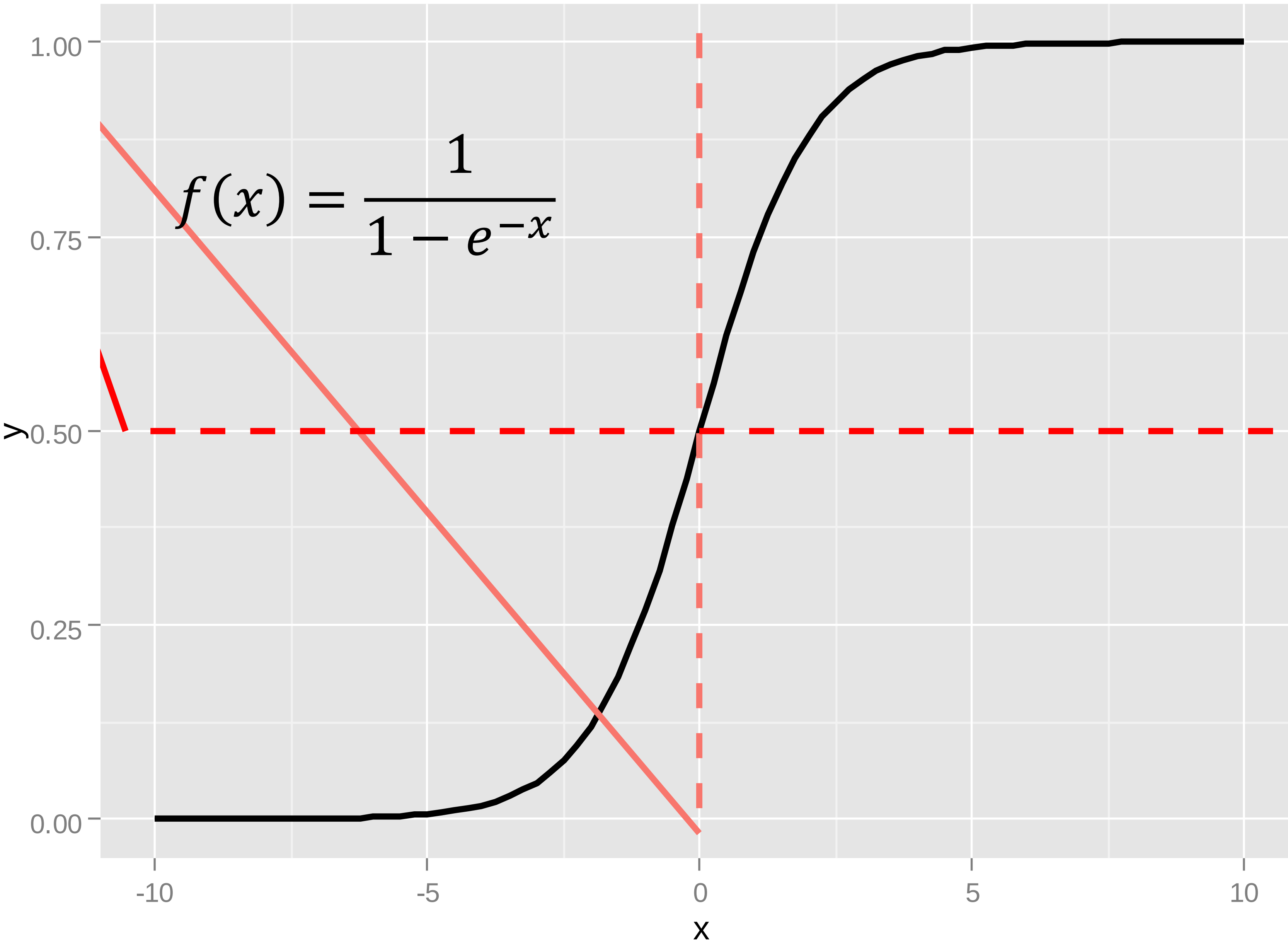
Inputs ? Outputs

`(-Inf, Inf)` → [0,1]

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

Logistic function

$$P(y) \sim \hat{y} = \frac{1}{1 + e^{-x}}$$



$f(x) = \frac{1}{1 - e^{-x}}$
1.00
0.75
0.50
0.25
0.00
y
-10
-5
0
5
10
x

Now What

- Proceed as we would with linear regression … and look for **β**'s

$$\hat{y} \sim \frac{1}{1+e^{-x}}$$

$$\hat{y} \sim \frac{1}{1+e^{-\beta_0+\sum_{i=1}^{p}\beta_i x_i}}$$

- Then solve as linear regression:

$$argmin_{\boldsymbol{\beta}}\left(\sum(\hat{y}-y)^2\right)$$

Not Done

- How do you go from [0,1] back to our binomial categories?

- Choice is somewhat arbitrary
  - `P=0.5`
  - Calibrate response

- Often don't care … you are interested in the probability anyway.

Worked Example

Exercise

- Write a the inverse logistic function ....
  - Not for Credit … No peeking on the web.

APPENDIX

Comprehensive ML Process