DECISION PATTERNS

Logistic Regression

Practical Machine Learning (with R)

UC Berkeley

Fall 2016

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!

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

  - Problem …

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

FUNCTION …

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

"Looks Math-y"

Need a tool …

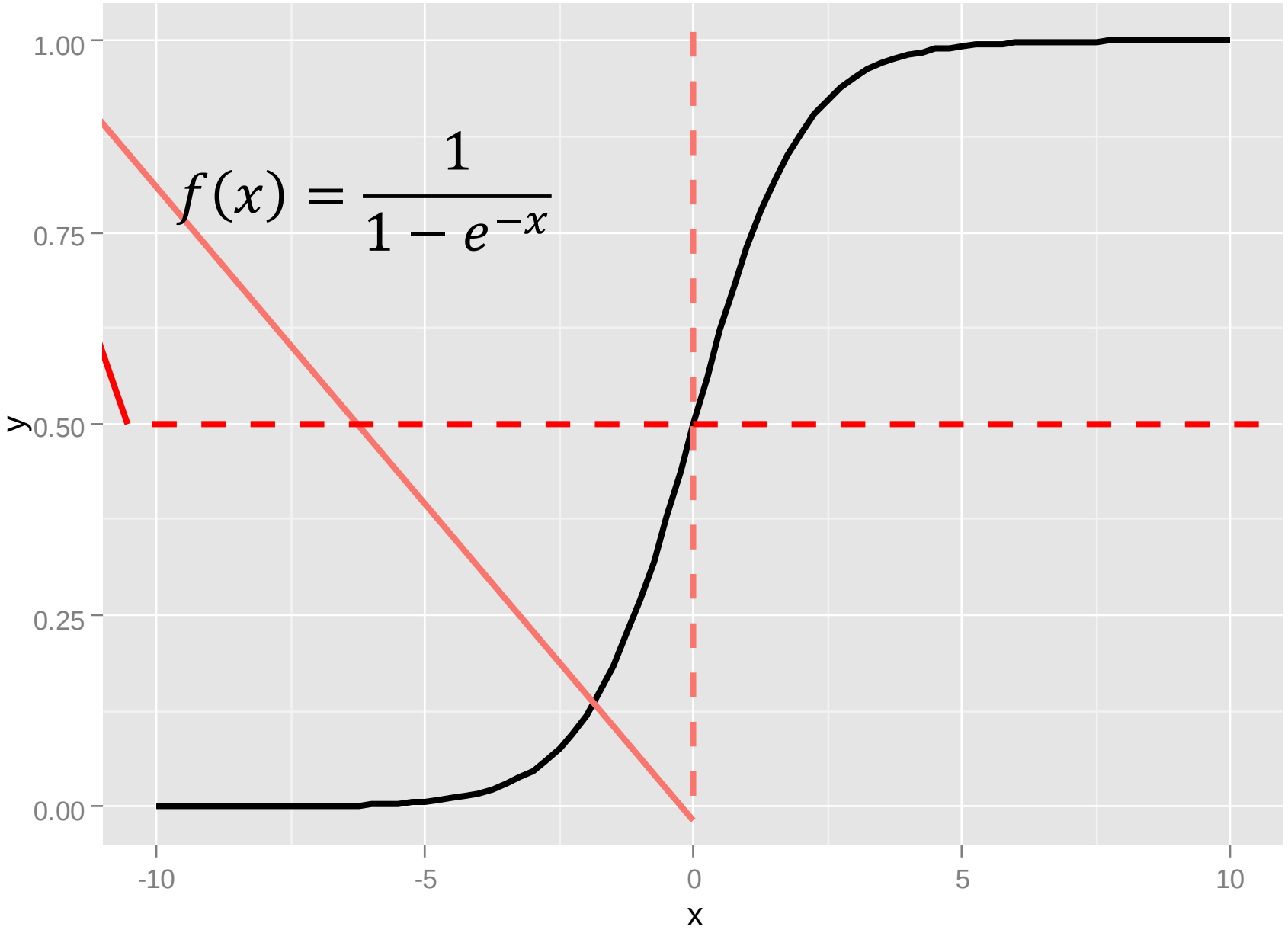
Inputs ? Outputs

`(-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: NYC Flights

APPENDIX

Worked Example: GermanCredit