

Logistic Regression

Akshata Mohan, Valerie Roth
MSBA, Class of 2017

References

<http://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf>

Previously ...

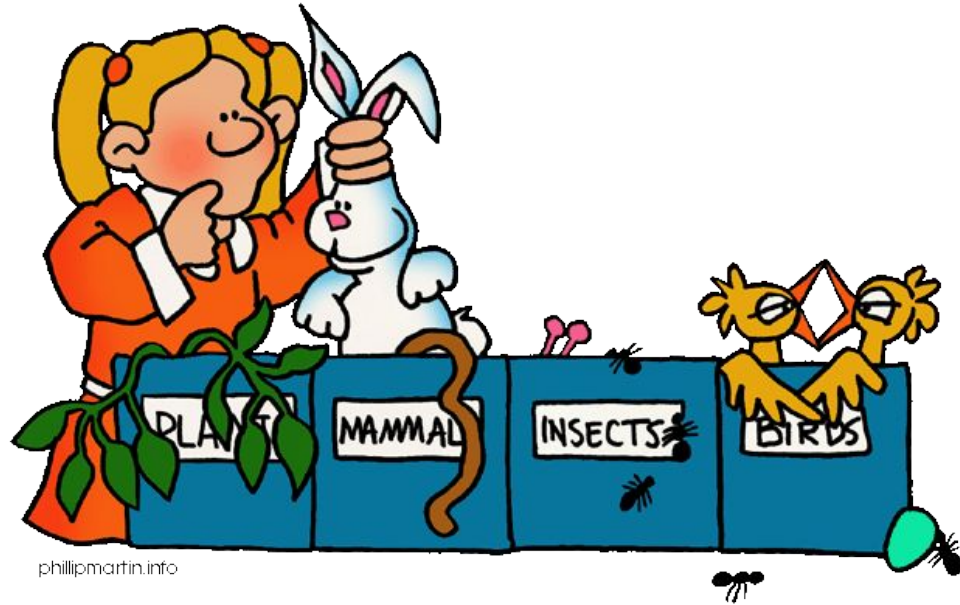
- Predicting continuous outcomes - conditioning Y on some values of X - conditional expectations
- Not enough sometimes

Why wasn't it enough?

- Discrete outcomes
- Assigning class labels to data - classification
- Basic model - “yes” or “no”
- Model to take noise into account - stochastic model - probabilities
- Conditional distribution of Y given input X
- $P(Y|X)$ - measure “precision” of prediction

Bottom line -

CLASSIFICATION!



Secret sauce

$$\log \frac{p(x)}{1 - p(x)} = \beta_0 + x \cdot \beta$$

How does this make sense?

- Linear dependence but 0,1 as o/p
- Sigmoid function [0,1]

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

Decision boundary

- Find B's
- Plug into equations
- $Eq \geq 0$ if $Y = 1$
- Figure represented is decision boundary
- Derivation of line, plane, non linear decision boundaries

Logistic regression from scratch

Logistic regression using sklearn

Measuring performance

- Accuracy
- ROC - receiver operating characteristic curves

ROC curves

