

DAT1

Feb 13th 2016

Agenda

- Review of last lessons
- Naive Bayes
- Clustering

LOGISTIC REGRESSION

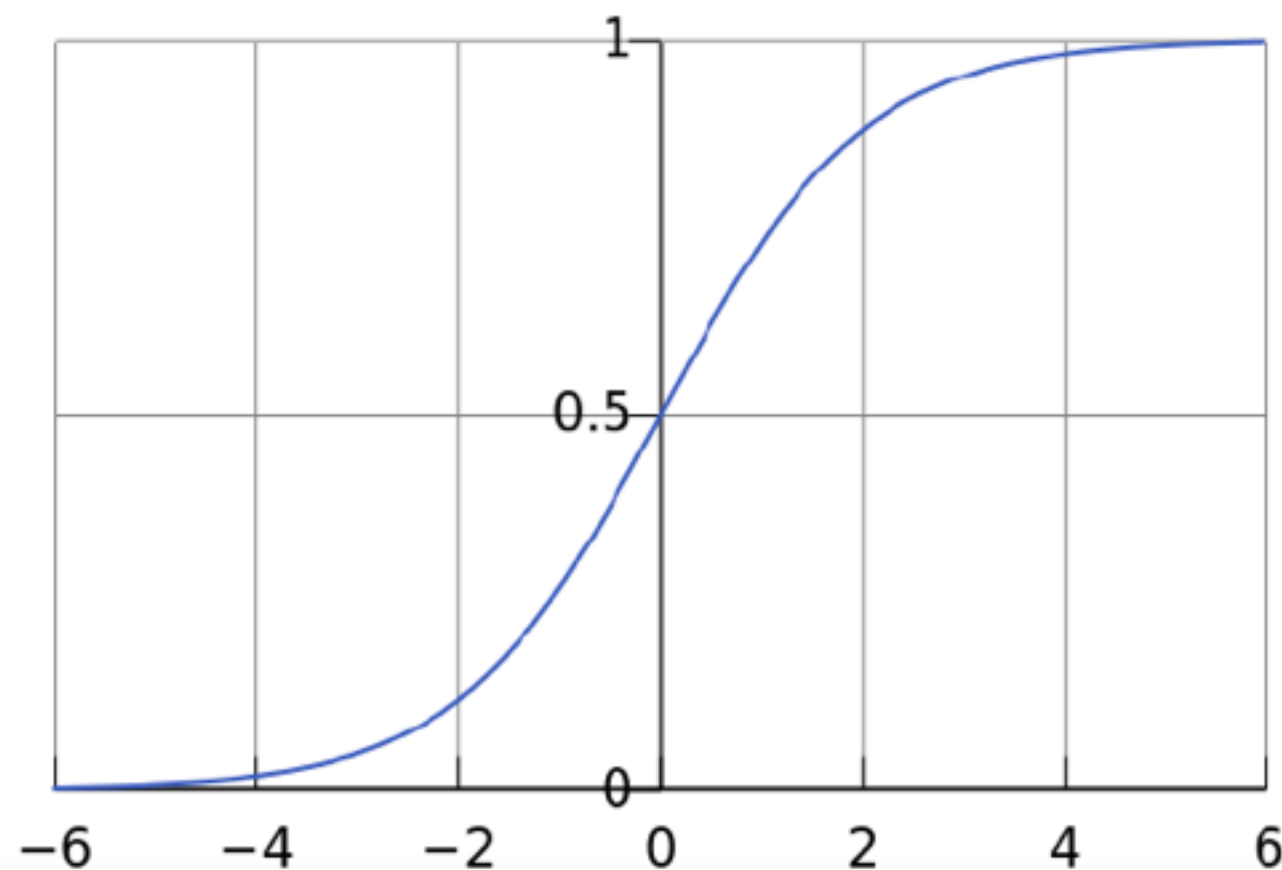
Q: What is logistic regression?

A: A generalization of the linear regression model to *classification* problems.

THE LOGISTIC FUNCTION

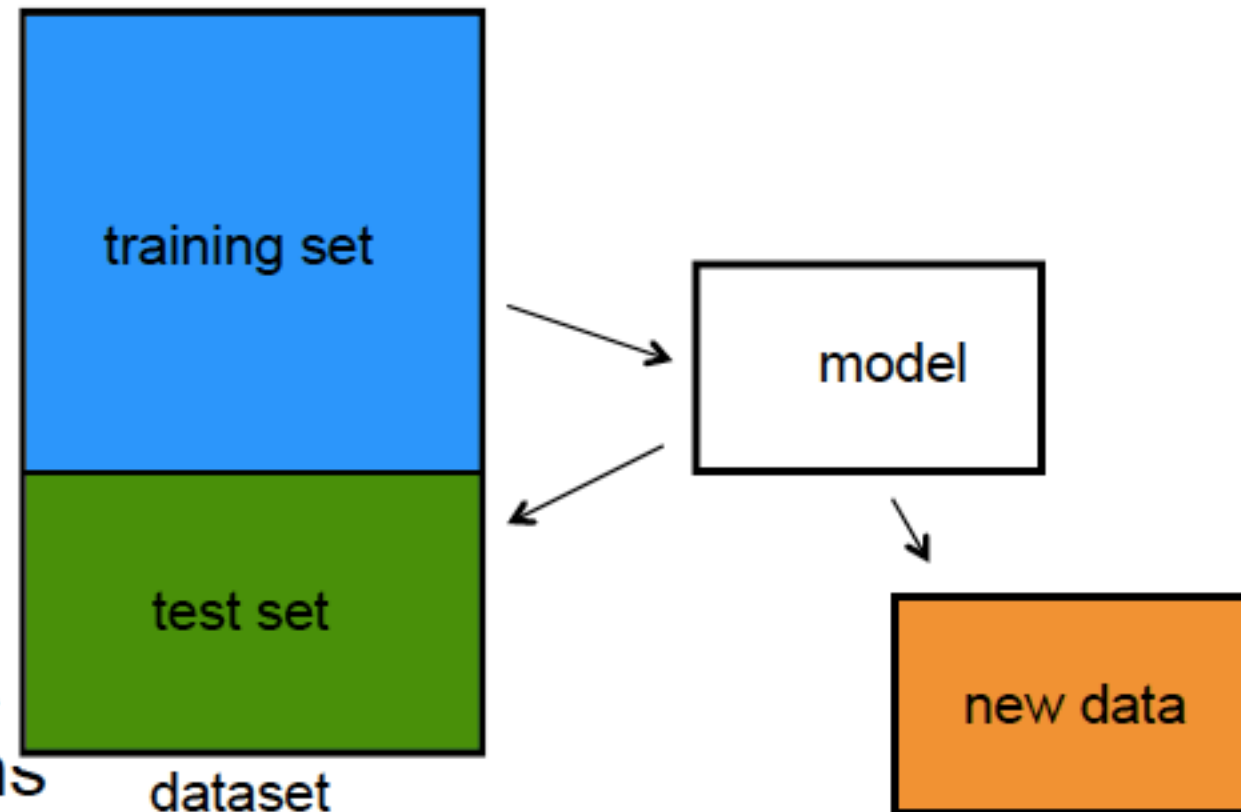
- The logistic function always returns a value between zero and one.

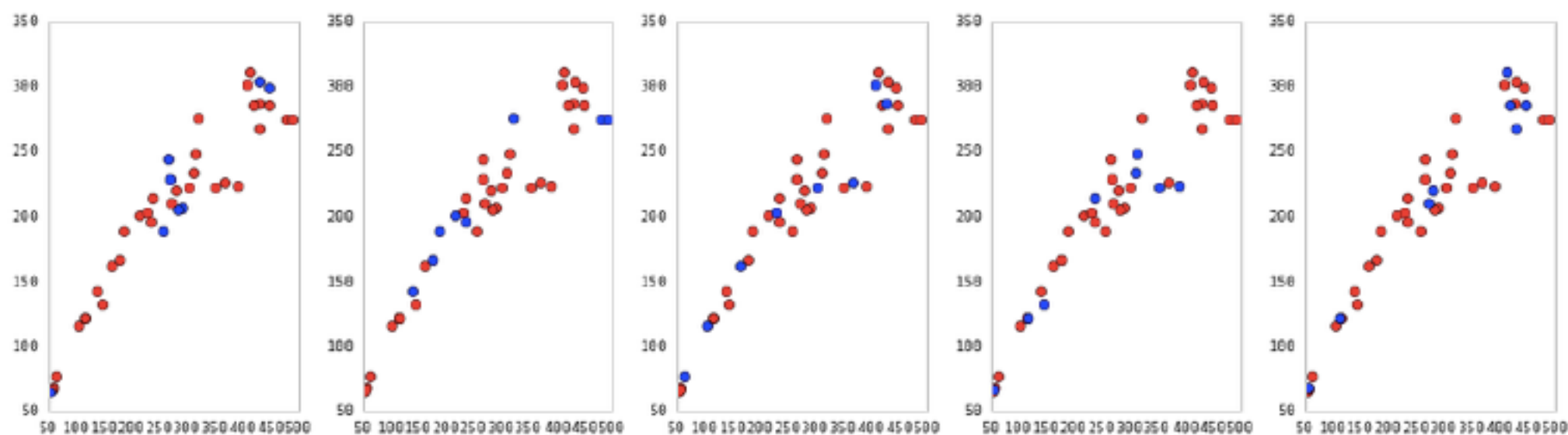
$$F(t) = \frac{1}{1 + e^{-t}}$$



Q: How can we make a model that generalizes well?

- 1) split dataset
- 2) train model
- 3) test model
- 4) parameter tuning
- 5) choose best model
- 6) train on **all** data
- 7) make predictions on new data





5-fold cross-validation: red = training folds,
blue = test fold

Source: http://nbviewer.ipython.org/github/fonnesbeck/Bios366/blob/master/notebooks/Section6_3-Model-Selection-and-Validation.ipynb