DATA 8

Fall 2016

Lecture 38, November 30

Updating Predictions Based on New Data

Slides created by Ani Adhikari and John DeNero

Announcements

- Nothing more is due!!!
- We will post some practice problems on this week's material, but they will not be due.
- Schedule for RRR week will be announced on Friday.

k Nearest Neighbor Classifier

- To classify a new point:
  - Based on the classes of the *k* nearest neighbors,
  - pick the class that is “more likely than not”

(Demo)

“More Likely Than Not” Game

- I give you some data and a point.
- You classify that new point as Class A or Class B, choosing the class that is “more likely than not”.

Round One

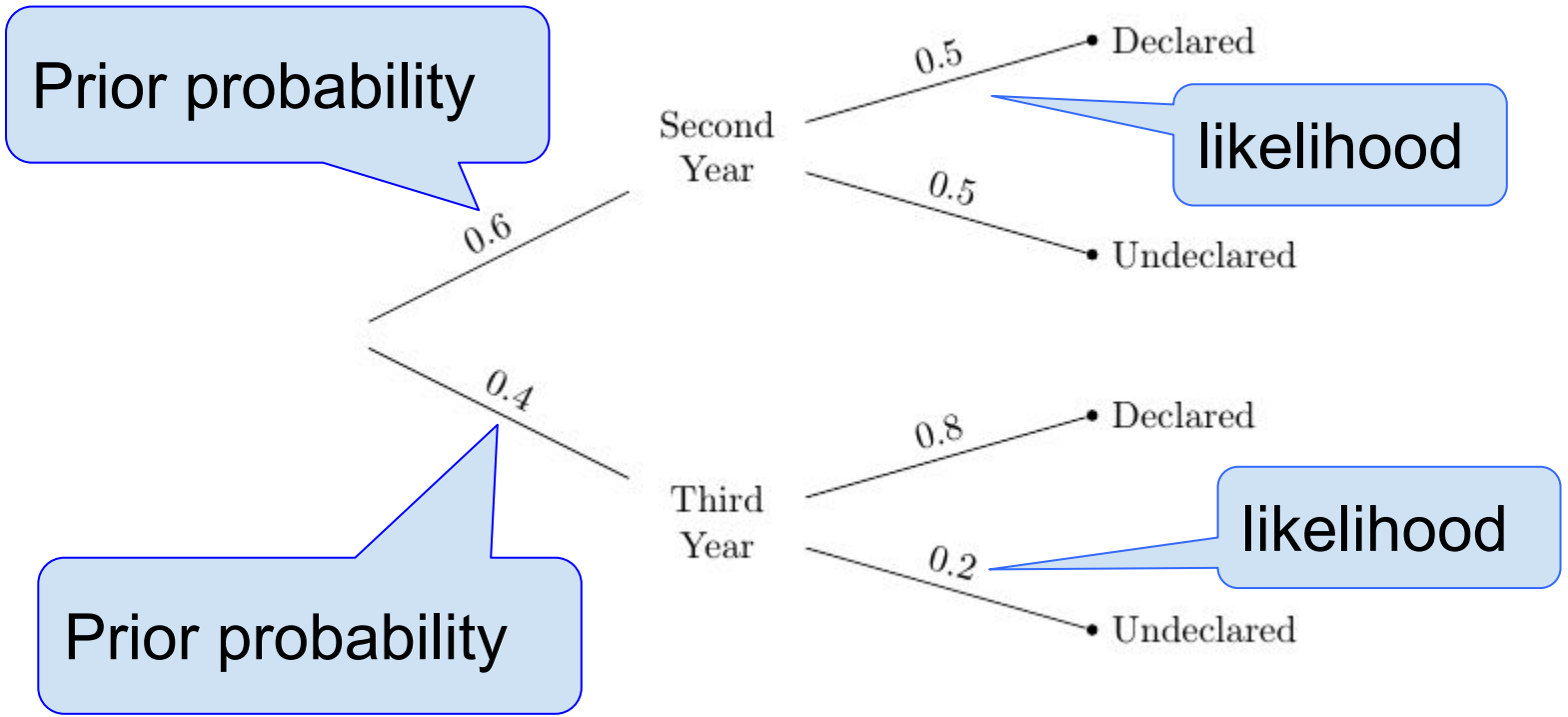
- Data:
  - Class consists of second years (60%) and third years (40%)
  - 50% of the Second Years have declared their major
  - 80% of the Third Years have declared their major
  - I pick one student at random.
- Second Year or Third Year?
  - Classify as Second Year because that's "more likely than not": chance = 60%

Round Two

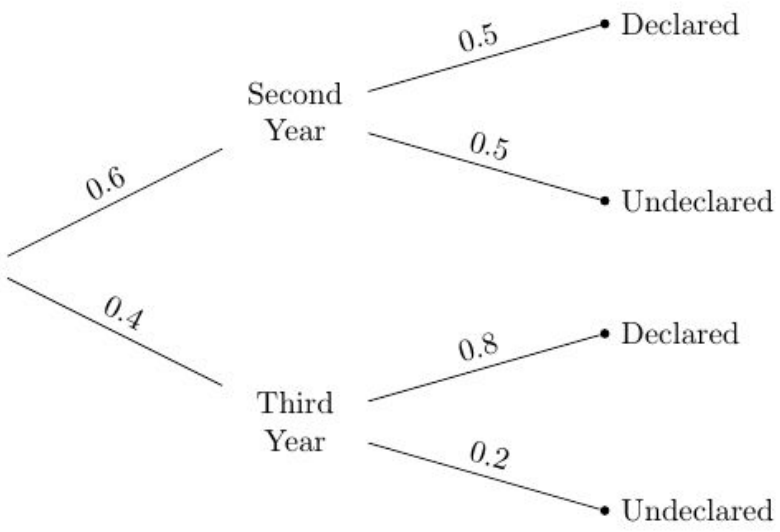
- Same data, almost ...
  - Class consists of second years (60%) and third years (40%)
  - 50% of the Second Years have declared their major
  - 80% of the Third Years have declared their major
  - I pick one student at random. **The student has declared a major!**
- Second Year or Third Year?

(Demo)

Terminology



Bayes' Rule



Pick a student at random.

Posterior probability:

P(Third Year | Declared)

$$= \frac{0.4 \times 0.8}{(0.6 \times 0.5) + (0.4 \times 0.8)}$$

= 0.5161…

Purpose of Bayes' Rule

- Update your prediction based on new information

- In a multi-stage experiment, find the chance of an event at an earlier stage, given the result of a later stage

(Demo)