Statistical Inference

Toolbox so far

- Uninformed search
- Heuristic search
- Local search
- Constraint satisfaction problems
- Probability and Bayes nets

Bayesian networks (Bayes nets)

- Specify a full joint probability distribution.
  - Uses conditional and marginal independences to represent information compactly.
  - Example of a **probabilistic model**.
- All probability questions have a **unique right answer.**
  - We can use the exact inference algorithm for Bayes nets to find it.

Real world

- Real world situations are often missing a model.
- We only have a small handful of observations about the world and we aren't 100% sure about how things relate to each other.
- How can we make probability estimates now?

Statistical inference

- Statistical inference lets us make probability estimations from observations about the way the world works, even if those observations don't tell the full story.
  - How likely is this email spam?
  - What is the probability it will rain tomorrow?
  - If I visit a certain house when trick-or-treating, what is the chance I'll get a Snickers bar?

Types of inference

- Hypothesis testing:
  - Given two or more hypotheses (events), decide which one is more likely to be true based on some data.
  - Example: Is this email spam or not spam?
- Parameter inference:
  - Given a model that is missing some probabilities, estimate those probabilities from data.
  - Example: Estimate bias of a coin from flips.

Hypothesis testing

- Let D be the event that we have observed some data.
  - Ex: D = received an email containing "cash" and "viagra"
- Let H_1, ..., H_k be disjoint, exhaustive events representing hypotheses to choose between.
  - Ex: H_1 = this email is spam, H_2 = it's not spam.
- How do we use D to decide which H is most likely?

Maximum likelihood

- Suppose we know or can estimate the probability $P(D \mid H_i)$ for each H_i.
- The maximum likelihood hypothesis is:

$$H^{ML} = \arg\max_i P(D \mid H_i)$$

- How to use it: compute $P(D \mid H_i)$ for each hypothesis and select the one with the greatest value.

- There are 2 boxes of cookies. One contains half chocolate chip cookies and half oatmeal raisin cookies. The other contains one third chocolate chip cookies and two thirds oatmeal raisin. I select a box and pull a random cookie from it. You observe that the cookie is chocolate chip. Which box is most likely to be the one that I chose from?

- I know that when my parents send me a check, there is an 98% chance that they will send it in a yellow envelope. I also know that when my dentist sends me a bill, there is a 5% chance that she will send it in a yellow envelope. Suppose a yellow envelope arrives on my doorstep. What is the maximum likelihood hypothesis regarding the sender?

Why ML sometimes is bad

- Suppose I tell you that there is a 3% chance that my any given envelope will be from my parents and a 97% chance that any given envelope will be from my dentist. Does it still seem likely that the envelope contains a check from my parents?

Bayesian reasoning

- Rather than compute $P(D \mid H_i)$, let's compute $P(H_i \mid D)$.
- What is the posterior probability of H_i given D?

$$P(H_i \mid D) = \frac{P(D \mid H_i)P(H_i)}{P(D)} = \alpha P(D \mid H_i)P(H_i)$$

MAP hypothesis

- Maximum a posteriori hypothesis is the Hi that maximizes the posterior probability:

$$H^{ML} = \text{argmax}_i P(H_i | D)$$

$$H^{ML} = \text{argmax}_i \frac{P(D | H_i)P(H_i)}{P(D)}$$

$$H^{ML} = \text{argmax}_i P(D | H_i)P(H_i)$$

ML vs MAP

$$H^{ML} = \arg\max_i P(D \mid H_i)$$

$$H^{MAP} = \arg\max_i P(D \mid H_i) P(H_i)$$

- The MAP hypothesis takes the prior probability of each hypothesis into account, ML does not.

- There are 2 boxes of cookies. One contains half chocolate chip cookies and half oatmeal raisin cookies. The other contains one third chocolate chip cookies and two thirds oatmeal raisin. I select a box and pull a random cookie from it. You observe that the cookie is chocolate chip. If you know that there’s a 90% chance that box 1 is on the table, while there’s only a 10% chance that box 2 is on the table, which box is most likely to be the one that I chose from?

- I know that when my parents send me a check, there is an 98% chance that they will send it in a yellow envelope. I know that when my dentist sends me a bill, there is a 5% chance that she will send it in a yellow envelope. Unfortunately, I also know that there is a only a 3% chance that any given envelope will be from my parents, while there is a is a 97% chance that any given envelope will be from my dentist. Suppose a yellow envelope arrives on my doorstep. What is the MAP hypothesis regarding the sender?

- There are 3 robots. Robot 1 will hand you a snack drawn at random from 2 doughnuts and 7 carrots. Robot 2 will hand you a snack drawn at random from 4 apples and 3 carrots. The third will hand you a snack drawn at random from 7 burgers and 7 carrots. Suppose you approach a robot and then eat the snack it hands you. If you eat a carrot, is it more likely that you approached robot 1 or 3? What if the prior probability of you approaching robot 1 is 70%, the prior probability of you approaching robot 2 is 20%, and the prior probability of you approaching the third robot is 10%?

ML vs MAP

$$H^{ML} = \arg\max_i P(D \mid H_i)$$

$$H^{MAP} = \arg\max_i P(D \mid H_i) P(H_i)$$

- When are the two hypothesis predictions the same?

Probability vs hypothesis

- Sometimes you only care about which hypothesis is more likely, and sometimes you need the actual probability.
- With ML, you have the probability already.
- With MAP, you need an extra step.

$$P(H_i \mid D) = \frac{P(D \mid H_i)P(H_i)}{P(D)} = \frac{P(D \mid H_i)P(H_i)}{\sum_i P(D \mid H_i)P(H_i)}$$