Statistical Inference

Toolbox so far

- Uninformed search
  - BFS, DFS, Dijkstra's algorithm (Uniform-cost search)
- Heuristic search
  - A*, greedy best-first search
- Probability and Bayes nets
  - Exact inference algorithm, approximate inference algorithms

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 (ML) 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.

- Two of my friends, Alice and Bob, bring cookies to the office!
- Alice's has baked an equal number of both chocolate chip and oatmeal raisin cookies.
- Bob's has baked chocolate chip and oatmeal raisin and as well, but twice as many oatmeal raisin as chocolate chip.
- I ask my friend to get me a cookie; they come back with a chocolate chip one.
- Is my cookie more likely to have been baked by Alice or Bob?

- 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 | H_i), let's compute P(H_i | 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 (MAP) hypothesis*** is the H_i that maximizes the posterior probability:

$$H^{MAP} = \text{argmax}_i \, P(H_i \mid D)$$

$$H^{MAP} = \text{argmax}_i \, \frac{P(D \mid H_i)P(H_i)}{P(D)}$$

$$H^{MAP} = \text{argmax}_i \, P(D \mid H_i)P(H_i)$$

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.

- Two of my friends, Alice and Bob, bring cookies to the office!
- Alice's has baked an equal number of both chocolate chip and oatmeal raisin cookies.
- Bob's has baked chocolate chip and oatmeal raisin and as well, but twice as many oatmeal raisin as chocolate chip.
- I ask my friend to get me a cookie. **Suppose I know that my friend picks Alice's cookies 90% of the time.** My friend comes back with a chocolate chip one.
- Is my cookie more likely to have been baked by Alice or Bob?

- 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.
- Robot 3 will hand you a snack drawn at random from 7 burgers and 7 carrots.
- Suppose your friend goes up to a robot (you don't see this happen) and is given a carrot. Is it more likely that your friend approached robot 1 or 3?
- What if the prior probability of your friend approaching robots 1, 2, and 3 are 20%, 40%, and 40%, respectively?

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.

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D)}$$

$$= \frac{P(D \mid H_i)P(H_i)}{\sum_j P(D, H_j)}$$

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

Probability vs hypothesis

- In the robot problem, what is P(R3 | C)?

$$P(R_3|C) = \frac{P(C|R_3)P(R_3)}{P(C)}$$

$$P(R_3|C) = \frac{P(C|R_3)P(R_3)}{\sum_{i=1}^{3} P(C, R_i)}$$

$$P(R_3|C) = \frac{P(C|R_3)P(R_3)}{\sum_{i=1}^{3} P(C|R_i)P(R_i)}$$

= (7/9 * 2/10) / (7/9 * 2/10 + 3/7 * 4/10 + 1/2 * 4/10) =~ 0.2952