Reminders

- Check the course website: http://www.cs.umass.edu/~wallach/courses/s12/cmpsci240/
- No homework this week
Inference
So far we have assumed the existence of a fully-specified probabilistic model that obeys the axioms of probability.
So far we have assumed the existence of a fully-specified probabilistic model that obeys the axioms of probability.

The questions we asked have had a unique right answer with respect to that model, e.g., if a fair die is rolled three times, what is the probability that all three rolls are greater than 3?
Statistical Inference

- In statistical inference, we are given only observations.
In statistical inference, we are given only observations

There may not be a single “right” answer, e.g., based on these emails, how likely is it that this new email is spam?
Statistical Inference

- In statistical inference, we are given only observations.
- There may not be a single “right” answer, e.g., based on these emails, how likely is it that this new email is spam?
- For the next few classes we will discuss different methods and techniques that can be used to answer questions like this.
Types of Inference

- **Hypothesis testing**: given two or more hypotheses, decide which one is more likely to be true based on some data, e.g., determine whether an email containing a particular set of words is more likely to be spam or genuine?
Types of Inference

- **Hypothesis testing**: given two or more hypotheses, decide which one is more likely to be true based on some data, e.g., determine whether an email containing a particular set of words is more likely to be spam or genuine?

- **Parameter inference**: have model that is fully specified except for some unknown parameters that we need to estimate, e.g., estimate the bias of a coin from a sequence of coin flips.
Let $D$ be the event that we have observed some data, e.g., $D = \text{observed an email containing “ca$h” and “viagra”}$
Hypothesis Testing

- Let $D$ be the event that we have observed some data, e.g., $D = \text{observed an email containing “ca$h” and “viagra”}$
- Let $H_1, \ldots, H_k$ be disjoint, exhaustive events representing hypotheses that we want to choose between, e.g., $H_1 = \text{event that email is spam}$, $H_2 = \text{event that email is not spam}$
Let $D$ be the event that we have observed some data, e.g., $D = \text{observed an email containing “ca$h” and “viagra”}$.

Let $H_1, \ldots, H_k$ be disjoint, exhaustive events representing hypotheses that we want to choose between, e.g., $H_1 = \text{event that email is spam}$, $H_2 = \text{event that email is not spam}$.

How do we use $D$ to decide which hypothesis is most likely?
Maximum Likelihood

- Suppose that we know or can compute the probability $P(D \mid H_i)$ of event $D$ given each hypothesis $H_i$
Maximum Likelihood

- Suppose that we know or can compute the probability $P(D \mid H_i)$ of event $D$ given each hypothesis $H_i$.

- The **maximum likelihood hypothesis** is the hypothesis that assigns the highest probability to the observed data:

$$H_{ML} = \arg\max_i P(D \mid H_i)$$
Maximum Likelihood

- Suppose that we know or can compute the probability $P(D \mid H_i)$ of event $D$ given each hypothesis $H_i$

- The maximum likelihood hypothesis is the hypothesis that assigns the highest probability to the observed data:

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

- How to use it: compute $P(D \mid H_i)$ for $i = 1 \ldots k$ hypotheses and select the hypothesis with the greatest value
e.g., 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?
e.g., 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?
The Problem with Maximum Likelihood

▶ e.g., 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 computing $P(D \mid H_i)$ for each hypothesis and comparing these values, we can use $P(H_i \mid D)$ instead.
Bayesian Reasoning

- Rather than computing $P(D \mid H_i)$ for each hypothesis and comparing these values, we can use $P(H_i \mid D)$ instead.

- If we know $P(H_i)$ and $P(D \mid H_i)$ for $i = 1 \ldots k$ we can use Bayes’ rule and the law of total probability to compute the posterior probability of each hypothesis given $D$:

$$P(H_i \mid D) = \frac{P(D \mid H_i) P(H_i)}{P(D)} = \frac{P(D \mid H_i) P(H_i)}{\sum_i P(D \mid H_i) P(H_i)}$$
Properties of Bayesian Reasoning

- The posterior probability $P(H_i | D)$ is a refinement of our prior belief about hypothesis $H_i$ in light of the observed data $D$.
Properties of Bayesian Reasoning

- The posterior probability $P(H_i | D)$ is a refinement of our prior belief about hypothesis $H_i$ in light of the observed data $D$

- $P(H_i | D)$ increases with both $P(H_i)$ and $P(D | H_i)$
The posterior probability \( P(H_i \mid D) \) is a refinement of our prior belief about hypothesis \( H_i \) in light of the observed data \( D \).

- \( P(H_i \mid D) \) increases with both \( P(H_i) \) and \( P(D \mid H_i) \).
- \( P(H_i \mid D) \) decreases as \( P(D) \) increases: the more probable it is that the data will be observed in general, the less evidence \( D \) provides in support of any particular hypothesis \( H_i \).
What if we just want to choose the “best” hypothesis?

e.g., I have the result of some diagnostic test and I want to know if my computer has a virus; I don’t care about the probabilities, I just want to know if I should install Ubuntu
Maximum A Posteriori

- The maximum a posteriori hypothesis is the hypothesis that maximizes the posterior probability given $D$:

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

$$\propto \arg\max_i P(D \mid H_i) P(H_i)$$
The maximum a posteriori hypothesis is the hypothesis that maximizes the posterior probability given $D$:

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

$$\propto \arg\max_i P(D \mid H_i) P(H_i)$$

$P(D \mid H_i)$ is now weighted by the prior probability $P(H_i)$; unlikely hypotheses are therefore downweighted accordingly.
Examples of Maximum a Posteriori

- e.g., 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 2 is on the table, while there’s only a 10% change that box 2 is on the table, which box is most likely to be the one that I chose from?
Examples of Maximum a Posteriori

- e.g., 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 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?
Examples of Maximum a Posteriori

- e.g., there are 3 robots. Robot 1 will hand you a snack drawn at random from 2 KFC Double Downs and 7 carrots. Robot 2 will hand you a snack drawn at random from 4 cupcakes 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%?
Comparing ML and MAP

- The ML hypothesis is the hypothesis that assigns the highest probability to the data: \( H_{\text{ML}} = \arg\max_i P(D \mid H_i) \)
Comparing ML and MAP

- The ML hypothesis is the hypothesis that assigns the highest probability to the data: $H^{ML} = \arg\max_i P(D \mid H_i)$

- The MAP hypothesis maximizes the posterior probability: $H^{MAP} = \arg\max_i P(H_i \mid D) = \arg\max_i P(D \mid H_i) P(H_i)$
Comparing ML and MAP

- The ML hypothesis is the hypothesis that assigns the highest probability to the data: $H_{ML} = \arg\max_i P(D \mid H_i)$

- The MAP hypothesis maximizes the posterior probability: $H_{MAP} = \arg\max_i P(H_i \mid D) = \arg\max_i P(D \mid H_i) P(H_i)$

- When are ML and MAP the same?
For Next Time

- Check the course website: http://www.cs.umass.edu/~wallach/courses/s12/cmpsci240/