More Statistical Inference

Review

- Let event D = data we have observed.
- Let events $H_1, \ldots, H_k$ be events representing hypotheses we want to choose between.
- Use D to pick the "best" H.

- There are two "standard" ways to do this, depending on what information we have available.

Maximum likelihood hypothesis

- The maximum likelihood hypothesis (H^{ML}) is the hypothesis that maximizes the probability of the data given that hypothesis.

$$H^{\mathrm{ML}} = \underset{i}{\operatorname{argmax}}\, P(D \mid H_i)$$

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

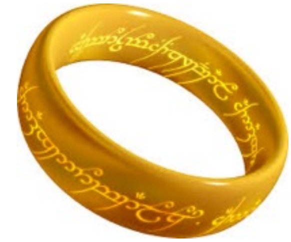
Maximum a posteriori (MAP) hypothesis

- The MAP hypothesis is the hypothesis that maximizes the posterior probability:

$$
\begin{aligned}
H^{\mathrm{MAP}} &= \operatorname*{argmax}_i P(H_i \mid D) \\
&= \operatorname*{argmax}_i \frac{P(D \mid H_i) P(H_i)}{P(D)} \\
&\propto \operatorname*{argmax}_i P(D \mid H_i) P(H_i)
\end{aligned}
$$

- The $P(D \mid H_i)$ terms are now *weighted* by the hypothesis prior probabilities.

One slide to rule them all



- The maximum likelihood hypothesis is the hypothesis that maximizes the probability of the observed data:

$$H^{\mathrm{ML}} = \operatorname*{argmax}_i P(D \mid H_i)$$

- The MAP hypothesis is the hypothesis that maximizes the posterior probability given D:

$$H^{\mathrm{MAP}} = \operatorname*{argmax}_i P(D \mid H_i) P(H_i)$$

- $P(H_i)$ is called the prior probability (or just prior).
- $P(H_i|D)$ is called the posterior probability.

- 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. Which robot did your friend probably approach?
- What if the prior probability of your friend approaching robots 1, 2, and 3 are 20%, 40%, and 40%, respectively?

Probability vs hypothesis

- Sometimes you only care about which hypothesis is more likely, and sometimes you need the actual probability.

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

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

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

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

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

Probability vs hypothesis

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

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

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

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

= (1/2 * 4/10) / (7/9 * 2/10 + 3/7 * 4/10 + 1/2 * 4/10) =~ 0.3795

- Suppose I work in FJ in a windowless office. I want to know whether it's raining outside. The chance of rain is 70%. My colleague walks in wearing his raincoat. If it’s raining, there’s a 65% chance he’ll be wearing a raincoat. Since he’s very unfashionable, there’s a 45% chance he’ll be wearing his raincoat even if it’s not raining. My other colleague walks in with wet hair. When it’s raining there’s a 90% chance her hair will be wet. However, since she sometimes goes to the gym before work, there’s a 40% chance her hair will be wet even if it’s not raining.
- What’s the posterior probability that it’s raining?

- We can't solve this problem because we don’t have any information about the probability of Colleague 1 wearing a raincoat and Colleague 2 having wet hair occurring *simultaneously*.
- We don't know P(C, W | R).
- Let's make an *assumption* that C and W are conditionally independent given that it is raining (or not raining).
- P(C, W | R) = P(C | R) * P(W | R)
  - (and similarly for given ~R)

Combining evidence

- It is very common to make this independence assumption for multiple pieces of evidence (data).

$$P(H_i \mid D_1, \ldots, D_m) = \frac{P(D_1, \ldots, D_m \mid H_i) P(H_i)}{P(D_1, \ldots, D_m)}$$

$$= \frac{\big(P(D_1 \mid H_i) \cdots P(D_m \mid H_i)\big) P(H_i)}{P(D_1, \ldots, D_m)}$$

$$= \frac{\big(\prod_{j=1}^{m} P(D_j \mid H_i)\big) P(H_i)}{P(D_1, \ldots, D_m)}$$

where $P(D_1 \ldots, D_m) = \sum_{i=1}^{k} \Big(\prod_{j=1}^{m} P(D_j \mid H_i)\Big) P(H_i)$