MAP and ML Estimation

Vassilis Athitsos
CSE 4308/5360: Artificial Intelligence I
University of Texas at Arlington

Overview of Candy Bag Example

As described in slides by Russell and Norvig, for Chapter 20 of the 2^{nd} edition:

- Five kinds of bags of candies.
  - 10% are h_1: 100% cherry candies
  - 20% are h_2: 75% cherry candies + 25% lime candies
  - 40% are h_3: 50% cherry candies + 50% lime candies
  - 20% are h_4: 25% cherry candies + 75% lime candies
  - 10% are h_5: 100% lime candies
- Each bag has an infinite number of candies.
  - This way, the ratio of candy types does not change as we pick candies out of the bag.

Hypotheses and Observations

- We have a bag of a certain type, and we are picking candies out of that bag.
  - We don't know if our bag is of type 1, 2, 3, 4, or 5.
- We have five ***hypotheses***: h_1, h_2, h_3, h_4, h_5.
- Out of the bag, we pick T candies, whose types are: Q_1, Q_2, …, Q_T.
  - Each Q_j is equal to either C (cherry) or L ("lime").
  - These Q_j's are called the ***observations***.

Questions We Want to Answer

- What is $P(h_i \mid Q_1, \ldots, Q_t)$?
  - Probability of hypothesis i after t observations.
- What is $P(Q_{t+1} = C \mid Q_1, \ldots, Q_t)$?
  - Similarly, what is $P(Q_{t+1} = L \mid Q_1, \ldots, Q_t)$
  - Probability of observation t+1 after t observations.

Simplifying notation

- Define:
  - $P_t(h_i) = P(h_i \mid Q_1, \ldots, Q_t)$
  - $P_t(Q_{t+1} = C) = P(Q_{t+1} = C \mid Q_1, \ldots, Q_t)$?
- Special case: t = 0 (no observations):
  - $P_0(h_i) = P(h_i)$
    - $P_0(h_i)$ is the prior probability of h_i
  - $P_0(Q_1 = C) = P(Q_1 = C)$
    - $P_0(Q_1 = C)$ is the probability that the first observation is equal to C.

Questions We Want to Answer, Revisited

Using the simplified notation of the previous slide:

- What is $P_t(h_i)$?
  - Probability of hypothesis i after t observations.
- What is $P_t(Q_{t+1} = C)$?
  - Similarly, what is $P_t(Q_{t+1} = L)$
  - Probability of observation t+1 after t observations.

A Special Case of Bayes Rule

- In the solution, we will use the following special case of Bayes rule:
  - $P(A \mid B, C) = P(B \mid A, C) * P(A \mid C) / P(B \mid C)$.

Computing $P_t(h_i)$

- Let t be an integer between 1 and T:
- $P_t(h_i) = P(h_i \mid Q1, \ldots, Q_t) =$

$$\frac{P(Q_t \mid h_i, Q_1, \ldots, Q_{t-1}) * P(h_i \mid Q_1, \ldots, Q_{t-1})}{P(Q_t \mid Q_1, \ldots, Q_{t-1})} \Rightarrow$$

$$\Rightarrow \boxed{P_t(h_i) = \frac{P(Q_t \mid h_i) * P_{t-1}(h_i)}{P_{t-1}(Q_t)}}$$

Computing $P_t(h_i)$ (continued)

- The formula $$P_t(h_i) = \frac{P(Q_t \mid h_i) * P_{t-1}(h_i)}{P_{t-1}(Q_t)}$$

is recursive, as it requires

knowing $P_{t-1}(h_i)$. The base case is $P_0(h_i) = P(h_i)$.

- To compute $P_t(h_i)$ we also need $P_{t-1}(Q_t)$. We show how to compute that next.

Computing $P_{t+1}(Q_t)$

- $P_t(Q_{t+1}) = P(Q_{t+1} \mid Q_1, \ldots, Q_t) =$

$$\sum_{i=1}^{5} (P(Q_{t+1} \mid h_i)\, P(h_i \mid Q_1, \ldots, Q_t)) \Rightarrow$$

$$\boxed{P_t(Q_{t+1}) = \sum_{i=1}^{5} (P(Q_{t+1} \mid h_i)\, P_t(h_i))}$$

Computing $P_t(h_i)$ and $P_t(Q_{t+1})$

- Base case: t = 0.
  - $P_0(h_i) = P(h_i)$, where $P(h_i)$ is known.
  - $P_0(Q_1) = \sum_{i=1}^{5} (P(Q_1 \mid h_i) * P(h_i))$, where $P(Q_1 \mid h_i)$ is known.
- To compute $P_t(h_i)$ and $P_t(Q_{t+1})$:
- For j = 1, …, t
  - Compute $P_j(h_i) = \dfrac{P(Q_j \mid h_i) \;\; * \;\; P_{j-1}(h_i)}{P_{j-1}(Q_j)}$
  - Compute $P_j(Q_{j+1}) = \sum_{i=1}^{5} (P(Q_{j+1} \mid h_i) \; * \; P_j(h_i))$

Computing $P_t(h_i)$ and $P_t(Q_{t+1})$

- Base case: t = 0.
  - $P_0(h_i) = P(h_i)$, where $P(h_i)$ is known.
  - $P_0(Q_1) = \sum_{i=1}^{5} (P(Q_1 \mid h_i) * P(h_i))$, where $P(Q_1 \mid h_i)$ is known.
- To compute $P_t(h_i)$ and $P_t(Q_{t+1})$:
- For j = 1, …, t
  - Compute $P_j(h_i) = \dfrac{\overbrace{P(Q_j \mid h_i)}^{\text{known}} * \overbrace{P_{j-1}(h_i)}^{\text{computed at previous round}}}{\underbrace{P_{j-1}(Q_j)}_{\text{computed at previous round}}}$
  - Compute $P_j(Q_{j+1}) = \sum_{i=1}^{5} (\underbrace{P(Q_{j+1} \mid h_i)}_{\text{known}}) * \underbrace{P_j(h_i))}_{\text{computed at previous line}}$

MAP Estimate

- The Maximum a Posteriori (MAP) estimate answers the following question:
  - What is the most likely hypothesis given observations $Q_1, \ldots, Q_t$?
  - Answer:
    - Compute, for each h_i, $P_t(h_i)$, using the algorithm described in the previous slides.
    - Select the h_i with the highest value of $P_t(h_i)$. The winning h_i is called the *MAP estimate*.

ML Estimate

- The Maximum Likelihood (ML) estimate answers the following question:
  - What is the most likely hypothesis given observations $Q_1, \ldots, Q_t$?
  - Answer: Compute, for each h_i:
    - $P(Q_1, \ldots, Q_t \mid h_i) = P(Q_1 \mid h_i) * \ldots * P(Q_t \mid h_i)$
    - Select the h_i with the highest value of $P(Q_1, \ldots, Q_t \mid h_i)$. The winning h_i is called the *ML estimate*.

Comparison of MAP and ML

- When there is enough information to compute both the MAP estimate and the ML estimate, then the MAP is more accurate, because it uses more information.
  - The ML estimate ignores the prior probabilities $P(h_i)$. Alternatively we can think that the ML estimate assumes that all prior probabilities $P(h_i)$ are equal to each other.
- When the prior probabilities $P(h_i)$ are not known, then only the ML estimate can be computed.
- Computing the ML estimate is more simple than computing the MAP estimate.