CS3244 Tutorial 6

Wu Zhaoxuan
wu.zhaoxuan@u.nus.edu

2022.4.8

Next Tutorial 15/4/22 - Zoom!

- Vote
  - Tue 12/4/22 - 10AM
  - Wed 13/4/22 - 10AM

  - Tue 12/4/22 - 7PM
  - ***<u>Wed 13/4/22/ - 7PM (Confirmed)</u>***
- Submit your tutorial answers to the Luminus folder.
  - Tutorial 7 Submission (T11, T12, T13) → T13 (Name your file: "T13_A0123456Z_Name.pdf") Submission

Short Review for Bayesian Inference

- What is Bayesian inference?
    - Use Bayes' theorem to update the probability for a hypothesis as more evidence becomes available.

$$P(h \mid D) = \frac{P(D \mid h)p(h)}{P(D)}$$

- $P(h)$: prior belief of hypothesis h
- $P(D \mid h)$: likelihood of data D given h
- $P(D) = \sum_{h \in H} P(D \mid h)P(h)$: marginal likelihood/evidence of D
- $P(h \mid D)$: posterior belief of h given D

MAP and ML

- Maximum A Posteriori (MAP): the most probable hypothesis given the training data

$$
\begin{aligned}
h_{MAP} &= \arg\max_{h \in H} P(h \mid D) \\
&= \arg\max_{h \in H} \frac{P(D \mid h) P(h)}{P(D)} \\
&= \arg\max_{h \in H} P(D \mid h) P(h)
\end{aligned}
$$

- Maximum Likelihood (ML): the hypothesis with the highest likelihood $P(D \mid h)$

$$
h_{ML} = \arg\max_{h \in H} P(D \mid h)
$$

TM 6.3

Consider the concept learning algorithm FIND-G, which outputs a maximally general consistent hypothesis. Suppose that

$$P(D \mid h) = \begin{cases} 1 & \text{if } h \text{ is consistent with } D \\ 0 & \text{otherwise} \end{cases}$$

(a) Give a distribution for $P(h)$ under which FIND-G is guaranteed to output a MAP hypothesis.

TM 6.3

Consider the concept learning algorithm FIND-G, which outputs a maximally general consistent hypothesis. Suppose that

$$P(D \mid h) = \begin{cases} 1 & \textit{if } h \textit{ is consistent with } D \\ 0 & \textit{otherwise} \end{cases}$$

(a) Give a distribution for $P(h)$ under which FIND-G is guaranteed to output a MAP hypothesis.

Solution

Same as that in lecture sides page 10. Choose an (uninformative) uniform prior $P(h) = \dfrac{1}{|H|}$. Follow the steps Bryan explained during the lecture, you get

$$P(h \mid D) = \begin{cases} \dfrac{1}{|VS_{H.D}|} & \text{if } h \text{ is consistent with } D \\ 0 & \text{otherwise} \end{cases}$$

Every consistent hypothesis is a MAP hypothesis and FIND-G output a maximally general hypothesis which is consistent.

TM 6.3

Consider the concept learning algorithm FIND-G, which outputs a maximally general consistent hypothesis. Suppose that

$$P(D \mid h) = \begin{cases} 1 & \textit{if } h \textit{ is consistent with } D \\ 0 & \textit{otherwise} \end{cases}$$

(b) Give a distribution for $P(h)$ under which FIND-G is ***<u>NOT</u>*** *guaranteed to output a MAP hypothesis.*

TM 6.3

Solution

Idea: Make more general hypothesis less probable a priori.

$$\forall h, h' \in H \quad h >_g h' \rightarrow P(h) < P(h')$$

When h is inconsistent with D,

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)} = \frac{0 \times P(h)}{P(D)} = 0$$

When h is consistent with D,

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)} = \frac{1 \times P(h)}{P(D)} = \frac{P(h)}{P(D)}$$

Then, $\forall h, h' \in VS_{H,D} \quad h >_g h' \rightarrow P(h \mid D) < P(h' \mid D)$.

That is, more general consistent hypotheses favored by FIND-G are less probable a posteriori and therefore not guaranteed to be a MAP hypothesis.

TM 6.3

Consider the concept learning algorithm FIND-G, which outputs a maximally general consistent hypothesis. Suppose that

$$P(D\,|\,h) = \begin{cases} 1 & \textit{if } h \textit{ is consistent with } D \\ 0 & \textit{otherwise} \end{cases}$$

(c) Give a distribution for $P(h)$ under which FIND-G is guaranteed to output a ML hypothesis but not guaranteed to output a MAP hypothesis.

Solution

ML hypothesis is one that has maximum $P(D\,|\,h)$, which is 1 for all consistent hypotheses here. Therefore, every consistent hypothesis is a ML hypothesis and FIND-G is guaranteed to output it.

Just use the same $P(h)$ as part (b).

TM 6.1

Consider again the medical diagnosis example applying Bayes' rule, as described on page 6 of the "Bayesian Inference" lecture slides. Suppose that the doctor decides to order a second lab test for the same patient and the second test returns a positive result as well. What are the posterior probabilities of $cancer$ *and* $\neg cancer$ *following these two tests? Assume that the results of the two tests are conditionally independent given the patient's state of having cancer (or not).*

- The test returns a correct +ve result in only 98% of the cases in which cancer is actually present,
- and a correct –ve result in only 97% of the cases in which cancer is not present.
- Furthermore, 0.008 of the entire population have this cancer.

TM 6.1

Solution

$$P(cancer \mid ++) = \frac{P(++ \mid cancer)P(cancer)}{P(++)}$$

$$= \frac{P(+ \mid cancer)^2 P(cancer)}{P(+ \mid cancer)^2 P(cancer) + P(+ \mid \neg cancer)^2 P(\neg cancer)}$$

$$= 0.89589552238$$

$$P(\neg cancer \mid ++) = 0.10410447762$$

Or alternatively,

$$= \frac{P(+ \mid cancer)P(cancer \mid +)P(+)}{P(+ \mid cancer)P(cancer \mid +)P(+) + P(+ \mid \neg cancer)P(\neg cancer \mid +)P(+)}$$

$$= \frac{P(+ \mid cancer)P(cancer \mid +)}{P(+ \mid cancer)P(cancer \mid +) + P(+ \mid \neg cancer)P(\neg cancer \mid +)}$$

TM 6.4

In the analysis of concept learning on page 10 of the "Bayesian Inference" lecture slides, we assume that the input instances $\mathbf{x}_d$ *for* $d \in D$ *are fixed. Therefore, in deriving an expression for* $P(D \mid h)$*, we only need to consider the probability of observing the target outputs* $t_d = c(\mathbf{x}_d)$ *for* $d \in D$ *for these fixed input instances* $\mathbf{x}_d$ *for* $d \in D$*. Consider the more general setting in which the input instances are not fixed, but are drawn independently from some probability distribution defined over the instance space* X*. The data* D *must now be described as the set of ordered pairs* $\{\langle \mathbf{x}_d, t_d \rangle\}$ *and* $P(D \mid h)$ *must now reflect the probability of encountering the specific input instance* $\mathbf{x}_d$ *as well as the probability of the observed target output* t_d*. Show that the expression for the posterior belief* $P(h \mid D)$ *on page 10 of the "Bayesian Inference" lecture slides holds even under this more general setting.*

TM 6.4

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

Solution

$$P(D \mid h) = \prod_{d \in D} P(\mathbf{x}_d, t_d \mid h)$$

$$= \prod_{d \in D} P(t_d \mid h, \mathbf{x}_d) P(\mathbf{x}_d)$$

$$= \prod_{d \in D} P(t_d \mid h, \mathbf{x}_d) \prod_{d \in D} P(\mathbf{x}_d)$$

$$= \begin{cases} \prod_{d \in D} P(\mathbf{x}_d) & \text{if } h \text{ is consistent with } D, \\ 0 & \text{otherwise} \end{cases}$$

If h is inconsistent with D, then

$$P(h \mid D) = 0$$

If h is consistent with D, then

$$\sum_{h \in VS_{H,D}} P(h \mid D) = \sum_{h \in VS_{H,D}} \frac{\prod_{d \in D} P(\mathbf{x}_d)}{|H| P(D)}$$

$$1 = \frac{\prod_{d \in D} P(\mathbf{x}_d)}{|H| P(D)} \sum_{h \in VS_{H,D}} 1$$

$$P(D) = \prod_{d \in D} P(\mathbf{x}_d) \times \frac{|VS_{H,D}|}{|H|}$$

Now,

$$P(h \mid D) = \frac{\prod_{d \in D} P(\mathbf{x}_d) \times \frac{1}{|H|}}{\prod_{d \in D} P(\mathbf{x}_d) \times \frac{|VS_{H,D}|}{|H|}} = \frac{1}{|VS_{H,D}|}$$

BL 9

On page 5 of the lecture slides, it is said if $P(h) = P(h')$ for any $h, h' \in H$, then $h_{MAP} = h_{ML}$. Prove or disprove that if $h_{MAP} = h_{ML}$, then $P(h) = P(h')$ for any $h, h' \in H$.

Solution

Similar to the first tutorial question,

Suppose more specific hypotheses are more probable a priori,

$$\forall h, h' \in H \quad h >_g h' \rightarrow P(h) < P(h')$$

When h is inconsistent with D, we have $P(h \mid D) = 0$.

When h is consistent with D, $P(h \mid D) = \dfrac{P(D \mid h) \times P(h)}{P(D)} = \dfrac{P(h)}{P(D)}$.

Therefore, $\forall h, h' \in VS_{H,D} \quad h >_g h' \rightarrow P(h \mid D) < P(h' \mid D)$

In this case, the output of FIND-S is MAP, and it is also ML because the hypothesis is consistent. We have disproved the claim.

BL 10

Assume that 2% of the population in a country carry a particular virus. A test kit developed by a pharmaceutical firm is able to detect the presence of the virus from a patient's blood sample. The firm claims that the test kit has a high accuracy of detection in terms of the following conditional probabilities obtained from their quality control testing:

P(the kit shows positive | the patient is a carrier) = 0.998

P(the kit shows negative | the patient is not a carrier) = 0.996

(a) Given that a patient is tested to be positive using this kit, what is the posterior belief that he is not a carrier?

BL 10

(a) Given that a patient is tested to be positive using this kit, what is the posterior belief that he is not a carrier?

Let X and $\neg X$ to represent positive and negative tests. Let Y and $\neg Y$ to represent whether the patient is a carrier or not. We want $P(\neg Y | X)$.

We know $P(X | Y) = 0.998$, $P(\neg X | \neg Y) = 0.996$ and $P(Y) = 0.02$.

$$
\begin{aligned}
P(\neg Y | X) &= \frac{P(X | \neg Y)P(\neg Y)}{P(X | \neg Y)P(\neg Y) + P(X | Y)P(Y)} \\
&= \frac{0.004 \times 0.98}{0.004 \times 0.98 + 0.998 \times 0.02} \\
&= 0.164
\end{aligned}
$$

BL 10

Suppose that the patient doesn't entirely trust the result offered by the first kit (perhaps because it has expired) and decides to use another test kit. If the patient is again tested to be positive using this second kit, what is the (updated) likelihood that he is not a carrier? You can assume conditional independence between results of different test kits given the patient's state of virus contraction. Give your answer to 4 decimal places.

Solution

Same as TM 6.1,

$$P(\neg Y | X \times 2) = \frac{P(X|\neg Y)P(\neg Y|X)}{P(X|\neg Y)P(\neg Y|X) + P(X|Y)P(Y|X)}$$

$$= \frac{0.004 \times 0.164}{0.004 \times 0.164 + 0.998 \times 0.836}$$

$$= 0.0008$$

Thank you!

- Any questions?