ML INFOSEC
3: Probability Theory II

February 5, 2019

Conditional probability

Definition

If $P(B) > 0$, then

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$

is called the conditional probability of A, given B.

Lemma

If $0 < P(B) < 1$, *then*

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$$

Total probability

Theorem

Let A and $B_1, \ldots, B_m$ be events such that $B_i \cap B_j = \emptyset$ for $i \neq j$ and

$$\Omega = \bigcup_{i=1}^{m} B_i.$$

If $P(B_i) > 0$ for all i, then

$$P(A) = \sum_{i=1}^{m} P(A|B_i)P(B_i).$$

Bayes' theorem

Theorem

If $P(A), P(B) > 0$, *then*

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}.$$

Another interpretation of Bayes theorem

Update of the probability of a hypothesis, H, in light of some body of data, D:

$$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(D)}.$$

- $P(H)$ is the probability of the hypothesis before we see the data, called the **prior probability** or prior.
- $P(H|D)$ is the probability of the hypothesis **after** we see the data, called the **posterior**.
- $P(D|H)$ is the probability of the data under the hypothesis, called the **likelihood**.
- $P(D)$ is the probability of the data under any hypothesis, called the **normalising constant**.

Example: The Monty Hall problem (Let's Make a Deal)

- Monty shows you three closed doors and tells you that there is a prize behind each door: one prize is a car, the other two are blanks.The prizes are arranged at random.
- The object of the game is to guess which door has the car. If you guess right, you get the car.
- You pick a door, which we will call A. The other doors are B and C.
- Before opening the door you chose, Monty increases the suspense by opening either B or C, whichever does not have the car. (If the car is actually behind A, Monty can safely open B or C, so he chooses one at random.)
- Then Monty offers you the option to stick with your original choice or switch to the one remaining unopened door.

Should you stick or switch or does it make no difference?

Monty Hall II

Data D: Monty chooses door B

Three hypotheses: H_A, H_B, and H_C represent the hypothesis that the car is behind A, B, and C, respectively.

- The car is behind A: Monty can safely open B or C. So the probability that he chooses B is $1/2$.
- The car is behind B: Monty has to open door C, so the probability that he opens door B is 0.
- The car is behind C: Monty opens B with probability 1.

Hypothesis H	Prior $P(H)$	Likelihood $P(D\|H)$	$P(H) \times P(D\|H)$	Posterior $P(H\|D)$
H_A	$1/3$	$1/2$	$1/6$	$1/3$
H_B	$1/3$	0	0	0
H_C	$1/3$	1	$1/3$	$2/3$

Monty Hall III

Hypothesis H	Prior $P(H)$	Likelihood $P(D\|H)$	$P(H) \times P(D\|H)$	Posterior $P(H\|D)$
H_A	$1/3$	$1/2$	$1/6$	$1/3$
H_B	$1/3$	0	0	0
H_C	$1/3$	1	$1/3$	$2/3$

$$\begin{aligned} P(D) &= P(H_A) \cdot P(D|H_A) + P(H_B) \cdot P(D|H_B) + P(H_C) \cdot P(D|H_C) \\ &= 1/2 \end{aligned}$$

$P(H_A|D) = 1/3$ and $P(H_C|D) = 2/3$. The **maximum a posteriori** hypothesis is H_C.

Independent events

Definition

Two events A and B are independent if

$$P(A \cap B) = P(A)P(B).$$

Definition

Two events A and B are conditionally independent given an event C if

$$P(A \cap B | C) = P(A|C)P(B|C).$$

Random variables

Definition

Let (Ω, P) be a finite probability space and V be a set. A function $X : \Omega \to V$ is called a random variable. For $x \in V$ we write

$$\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$$

and

$$P(X = x) = P(\{X = x\}).$$

If $V \subset \mathbb{R}$, then X is called a real random variable. If $V \subset \mathbb{R}$ is finite, then

$$E(X) = \sum_{x \in V} xP(X = x)$$

is called the expectation of X.

Naive Bayes I

Let A_1, A_2, ..., A_k be finite sets of **attributes**, C a finite set of **classes** classes and $T \subset A_1 \times \cdots \times A_k$ a set of **instances**. Moreover, let

$$F : T \to C$$

a function, i.e. each instance $(x_1, ..., x_k)$ is classified as class $F(x_1, ..., x_k)$. How should we predict the class of a new instance $(a_1, ..., a_k) \notin T$?

Naive Bayes II

The idea is to choose a class c_{MAP} (MAP stands for **maximum a posteriori**) such that

$$P(c_{MAP}|a_1, ...a_k) = \max_{c \in C} P(c|a_1, \dots, a_k),$$

which can also be written as

$$c_{MAP} = \mathrm{argmax}_{c \in C} P(c|a_1, \dots, a_k).$$

By Bayes' theorem $P(c|a_1, ..., a_k)$ can be rewritten as

$$P(c|a_1, ..., a_k) = \frac{P(c)P(a_1, ..., a_k|c)}{P(a_1, ..., a_k)},$$

thus

Naive Bayes III

MAP classifier

For $a = (a_1, \dots, a_k)$, the Naive Bayes classifier is given by

$$c_{MAP} = \mathrm{argmax}_{c \in C} P(c) P(a_1, \dots, a_k | c)$$

$P(a_1, ..., a_k | c)$ and *a fortiori* c_{MAP} are difficult to calculate. Therefore, we make the following simplifying assumption:

Naive Bayes assumption

$a_1, ..., a_k$ are conditionally independent given the target class c:

$$P(a_1, ..., a_k | c) = \prod_{i=1}^{k} P(a_i | c).$$

Naive Bayes IV

Naive Bayes Classifier

For $a = (a_1, \ldots, a_k)$, the Naive Bayes classifier is given by

$$c_{NB}(a) = \mathrm{argmax}_{c \in C} P(c) \prod_{i=1}^{k} P(a_i | c).$$

Play tennis 1

$$\begin{array}{ll} A_1 = \{sunny, overcast, rainy\} & Outlook \\ A_2 = \{hot, mild, cool\} & Temperature \\ A_3 = \{high, normal\} & Humidity \\ A_4 = \{strong, weak\} & Wind \\ C = \{yes, no\} & Play\ \ tennis \end{array}$$

$$\begin{aligned} T \quad = \{ \quad & (sunny, hot, high, weak), (sunny, hot, high, strong), \\ & (overcast, hot, high, weak), (rainy, mild, high, weak), \\ & (rainy, cool, normal, weak), (rainy, cool, normal, strong), \\ & (overcast, cool, normal, strong), (sunny, mild, high, weak), \\ & (sunny, cool, normal, weak), (rainy, mild, normal, weak), \\ & (sunny, mild, normal, strong), (overcast, mild, high, strong), \\ & (overcast, hot, normal, weak), (rainy, mild, high, strong)\} \end{aligned}$$

Play tennis 2

Outlook a_1	Temperature a_2	Humidity a_3	Wind a_4	Class $c = F(a_1, \ldots, a_4)$
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

Play tennis 3

We wish to calculate the Naive Bayes classifier for $a = (sunny, cool, high, strong)$.

Play tennis	$P(yes) = 9/14$	$P(no) = 5/14$
Outlook	$P(sunny\|yes) = 2/9$	$P(sunny\|no) = 3/5$
Temperature	$P(cool\|yes) = 1/3$	$P(cool\|no) = 1/5$
Humidity	$P(high\|yes) = 1/3$	$P(high\|no) = 4/5$
Wind	$P(strong\|yes) = 1/3$	$P(strong\|no) = 3/5$
Product	$1/189 = 0.0053$	$18/875 = 0.0206$

$c_{NB}(sunny, cool, high, strong) = no$