Bayesian Classification

Exercise 1 : Probabilities

Show the relationship between the prior, posterior and likelihood probabilities.

Answer

The Theorem of Bayes gives:

$$P(A_i|B_1,...,B_p) = \frac{P(A_i) \cdot P(B_1,...,B_p|A_i)}{P(B_1,...,B_p)}$$

with the priors $P(A_i)$, the posteriors $P(A_i|B_1,...,B_p)$ and the likelihoods $P(B_1,...,B_p|A_i)$.

Exercise 2 : Application of Bayes Theorem

(adapted from Kashani 2021 "Deep Learning Interviews: Hundreds of fully solved job interview questions from a wide range of key topics in AI.")

The Dercum disease is an extremely rare disorder of multiple painful tissue growths. In a population in which the ratio of diabetics to non-diabetics is equal, 5% of diabetics and 0.25% of non-diabetics have the Dercum disease.

A person is chosen at random and that person has the Dercum disease. Calculate the probability that the person is diabetic.

Answer

$$\begin{aligned} P(\text{Dercum}|\text{diabetic}) &= 0.05 \\ P(\text{Dercum}|\text{non-diabetic}) &= 0.0025 \\ P(\text{non-diabetic}) = P(\text{diabetic}) &= 0.5 \end{aligned}$$

By Bayes Theorem, we get:

$$\begin{aligned} P(\text{diabetic}|\text{Dercum}) &= \frac{P(\text{diabetic}) \cdot P(\text{Dercum}|\text{diabetic})}{P(\text{diabetic}) \cdot P(\text{Dercum}|\text{diabetic}) + P(\text{non-diabetic}) \cdot P(\text{Dercum}|\text{non-diabetic})} \\ &= \frac{0.5 \cdot 0.05}{0.5 \cdot 0.05 + 0.5 \cdot 0.0025} \approx 0.9524 \end{aligned}$$

Exercise 3 : Problems of Naïve Bayes

Give at least two reasons why the results of a Naïve Bayes classifier may or may not be very good and which steps could be taken to influence them.

Answer

Possible reasons could be violation of NB assumption by (strongly) correlated features, extremely naive probability estimation, label noise, imbalanced classes, poor discretization, suboptimal feature scaling, . . .

Exercise 4 : Probability Basics (Kolmogorov)

Prove the implications of the Kolmogorov axioms from the lecture (Theorem 7).

Answer

Let Ω be a set, and let $A, B \in \mathcal{P}(\Omega)$. Then $P : \mathcal{P}(\Omega) \to \mathbf{R}$ is a probability measure if the following conditions hold:

(I) $P(A) \geq 0$

(II) $P(\Omega) = 1$

(III) $A \cap B = \emptyset$ implies $P(A \cup B) = P(A) + P(B)$

(1) To show: $P(A) + P(\overline{A}) = 1$.
From the axioms II and III follows: $P(\Omega) = 1 = P(A \cup \overline{A}) = P(A) + P(\overline{A})$

(2) To show: $P(\emptyset) = 0$.
Substitute Ω for A in (1).

(3) To show: $A \subseteq B$ implies $P(A) \leq P(B)$
Given $A \subseteq B$ the set B can be partitioned into two mutually exclusive events: $B = A \cup (\overline{A} \cap B)$. According to Axiom III we have $P(B) = P(A \cup (\overline{A} \cap B)) = P(A) + P(\overline{A} \cap B)$. According to Axiom I we have $P(\overline{A} \cap B) \geq 0$, which entails the claimed statement.

(4) To show: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Given $A \cup B = A \cup (\overline{A} \cap B)$ and $B = (A \cap B) \cup (\overline{A} \cap B)$. According to Axiom III we get $P(A \cup B) = P(A \cup (\overline{A} \cap B)) = P(A) + P(\overline{A} \cap B)$ and $P(B) = P(A \cap B) + P(\overline{A} \cap B)$, and it follows that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

	A	$\overline{A}$
B	$A \cap B$	$\overline{A} \cap B$
$\overline{B}$	$A \cap \overline{B}$	$\overline{A} \cap \overline{B}$

(5) To show: If $A_1, A_2, \ldots, A_n$ are mutually exclusive events then holds
$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$.
Proof by mathematical induction over n:

- Basic step. For n=2: there is nothing to prove (Axiom III).
- Induction hypothesis. If $A_1, \ldots, A_n$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$ holds.
- Inductive step. To show: The statement holds for $n+1$, too. Given A_{n+1} with $A_{n+1} \cap (A_1 \cup A_2 \cup \ldots \cup A_n) = \emptyset$. Then holds:

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n \cup A_{n+1}) &= P((A_1 \cup \ldots \cup A_n) \cup A_{n+1}) \\ &\overset{\text{(III)}}{=} P(A_1 \cup \ldots \cup A_n) + P(A_{n+1}) \\ &\overset{\text{Ind.Hyp.}}{=} P(A_1) + \ldots + P(A_n) + P(A_{n+1}) \end{aligned}$$