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 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, ...