CASE STUDY: PREDICTIVE MEDICAL DIAGNOSIS USING GRAPHICAL MODELS

In this case study, we will do a simple case study on using graphical models, including Bayesian networks and Naïve Bayes, for doing predictive medical diagnosis.

Problem statement and context

Medical diagnosis is the process of determining which disease or condition explains a person's symptoms and signs. It is most often referred to as diagnosis with the medical context being implicit. The information required for diagnosis is typically collected from a history and physical examination of the person seeking medical care. Often, one or more diagnostic procedures, such as medical tests, are also done during the process. Sometimes posthumous diagnosis is considered a kind of medical diagnosis.

Diagnosis is often challenging, because many signs and symptoms are nonspecific, and can only be undertaken by registered and licensed health professionals. For example, redness of the skin (erythema), by itself, is a sign of many disorders and thus does not tell the healthcare professional what is wrong. Thus differential diagnosis, in which several possible explanations are compared and contrasted, must be performed. This involves the correlation of various pieces of information followed by the recognition and differentiation of patterns.

Doctors can, however, make mistakes when diagnosing, especially for rare diseases or corner cases. In this case study, we explore how machine learning methods like Bayes nets or Naïve Bayes could be used for diagnosing a disease given some symptoms.

Technical Details

Patients see a doctor and complain about a number of symptoms (headache, 100F fever,…). What is the most likely disease d_i given the set of symptoms S the patient has?

$$\arg\max_{d_i} P\left(d_i \mid \overline{S}\right)$$

We will explore how we could use Naïve Bayes, and more generally, Bayesian networks, for addressing this medical diagnosis problem. First, we provide technical details on Naïve Bayes and on Maximum Likelihood estimation.

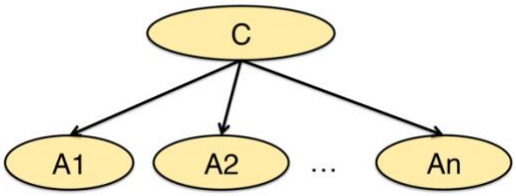
Naïve Bayes

Assume the items in your data set have a number of attribute $A_1 \ldots A_n$.

Each item also belongs to one of a number of given classes $C_1 \ldots C_k$.

Which attributes an item has depends on its class.

If you only observe the attributes of an item, can you predict the class?



$$\begin{aligned} \text{argmax}_C\, P(C \mid A_1 \ldots A_n) &= \\ &= \text{argmax}_C\, P(A_1 \ldots A_n \mid C)\, P(C) \\ &= \text{argmax}_C \textstyle\prod_j P(A_j \mid C)\, P(C) \end{aligned}$$

We need to estimate:

- the multinomial P(C)
- for each attribute A_j and class c $P(A_j \mid c)$

Maximum Likelihood estimation

If we have a set of training data where the class of each item is given:

- the multinomial P(C=c) = freq(c)/N
- for each attribute A_j and class c:
  $P(A_j = a \mid c) = \text{freq}(a, c)/\text{freq}(c)$

where

freq(c) = the number of items in the training data that have class c

freq(a, c) = the number of items in the training data that have attribute a and class c.

Case Questions

i) Consider the abstract Naïve Bayes example described in Technical Details. How would you model the disease prediction problem as a Naïve Bayes model? Try to be specific, using real diseases and symptoms if possible.

ii) What is the key assumption (or assumptions) in the Naïve Bayes model? What do you gain through such simplifications?

iii) Suppose the symptoms themselves were probabilistic (i.e. you are more sure about some symptoms than others). Can the Naïve Bayes handle such scenarios? If not, how would you extend the Naïve Bayes, perhaps by converting into a Bayes Net, to handle such eventualities?

iv) How would you use maximum likelihood to address the medical diagnosis problem? What is the training data and what are the classes?

v) Would you want to build a single graphical model for all diseases or want to separate out by disease type (or something else)?

vi) Go to a general medical website (like Mayo clinic) and pick a few diseases and symptoms. Try to build both a Bayes Net and Naïve Bayes for each of them. What structural assumptions are you making in your Bayes Net?