Naïve Bayes Classification

Problem: use a Naïve Bayes classifier to answer the question "Is today a good day to fish"?

Observations (training data):

Data	Wind	Water	Air	Forecast	Oracle
1	Strong	Warm	Warm	Sunny	Yes
2	Weak	Warm	Warm	Sunny	No
3	Strong	Warm	Warm	Cloudy	Yes
4	Strong	Moderate	Warm	Rainy	Yes
5	Strong	Cold	Cool	Rainy	No
6	Weak	Cold	Cool	Rainy	No
7	Weak	Cold	Cool	Sunny	No
8	Strong	Moderate	Warm	Sunny	Yes
9	Strong	Cold	Cool	Sunny	Yes
10	Strong	Moderate	Cool	Rainy	No
11	Weak	Moderate	Cool	Sunny	Yes
12	Weak	Moderate	Warm	Sunny	Yes
13	Strong	Warm	Cool	Sunny	Yes
14	Weak	Moderate	Warm	Rainy	No

Naïve Bayes classification

Learn a hypothesis based on estimated probabilities.

- Simplifying assumption: attribute values are independent, given the classification (e.g. Sunny forecast and Warm air are completely independent). So the probability of a specific instance (a conjunction of attribute values a_1, a_2,…a_i) is the product of the probability of each individual attribute, given the classification.

- To classify, assign the most probable class, given the observed attribute values:

$$C_{NB} = \max_{c_j \in C}(P(c_j)\prod_i P(a_i \mid c_j))$$

where $P(c_j)$ is the probability of class *j*, and $P(a_i|c_j)$ is the probability of attribute *i* given class j (conditional probability), estimated based on frequencies in the training data.

Example: new instance: <Strong Cold Warm Sunny> ⇨ Fish?

Learn probabilities based on training data:

1. Estimate the probability of each class by using its frequency in the training data. There are two classes, C = {Yes, No}. The probability of each class occurring is:

   P(Fish = Yes) = 8/14
   P(Fish = No) = 6/14

2. Estimate the probability of each specified attribute value, given a particular class (i.e. the conditional probabilities). There are four attributes, and two classes:

   P(Wind = Strong | Fish = Yes) = 6/8
   P(Wind = Strong | Fish = No) = 2/6

   P(Water = Cold | Fish = Yes) = 1/8
   P(Water = Cold | Fish = No) = 3/6

   P(Air = Warm | Fish = Yes) = 5/8
   P(Air = Warm | Fish = No) = 2/6

   P(Forecast = Sunny| Fish = Yes) = 6/8
   P(Forecast = Sunny | Fish = No) = 2/6

3. Compute the probability of each class, given the probabilities of the observed data.

   P(Yes) • P(Strong|Yes) • P(Cold|Yes) • P(Warm|Yes) • P(Sunny|Yes)
   = (0.57)(0.75)(0.125)(0.625)(0.75) = 0.025

   P(No) • P(Strong|No) • P(Cold|No) • P(Warm|No) • P(Sunny|No)
   = (0.43)(0.33)(0.5)(0.33)(0.33) = 0.008

Therefore, $\mathbf{C_{NB} = Yes}$