Naïve Bayes Classifier

Dr. Goutam Chakraborty

What Is Bayesian Analysis?

- *Bayesian analysis* is a field of statistics that is based on the notion of conditional probability.
- It can be viewed as the formalization of the process of incorporating scientific knowledge using probabilistic tools.
- It provides uncertainty quantification of parameters by its conditional distribution in the light of available data.

Bayes' Theorem

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

- P(A) is the prior probability of event A. It is called the *prior* because it does not take into account any information about event B.
- P(B|A) is the conditional probability of event B given event A.
- P(B) is the prior or marginal probability of event B.
- P(A|B) is the conditional probability of event A given event B. It is called the posterior probability because it is derived from the specified value of event B.

Bayesian Analysis

- The Bayesian approach to statistical inference treats parameters as random variables.
- It includes the incorporation of prior knowledge and its uncertainty in making inferences on unknown quantities (model parameters, missing data, and so on).
- It expresses the uncertainty concerning the parameter through probability statements and distributions.

Conceptual Basis of Naïve Bayes Method

- The Naive Bayes classifies observations into *classes* that are defined by the levels of a categorical target. The variables (or factors) that are used for classification are often called *features*
- For each class, the algorithm computes the conditional probability of each feature value occurring.
  - Classification is based on the idea that an observation whose feature values have high conditional probabilities within a certain class has a high probability of belonging to that class.
- For each new record to be classified;
  - Find all training records with the same predictor profile
  - Determine what classes the records belong to and which class is most common
  - Assign the most common class to the new record

Naïve Bayes Algorithm

- The Naive Bayes method classifies an observation into the class for which its probability of membership, given the values of its features, is highest.
  - The method assumes that the features are conditionally independent within each class
- Denote the possible classifications by $C_1, \ldots, C_K$.
- Denote the features, or predictors, by $X_1, X_2, \ldots, X_p$.
- The conditional probability that an observation with predictor values $x_1, x_2, \ldots, x_p$ belongs in the class C_k is computed as follows:

$$P(C_k|(x_1, \ldots, x_p)) = \frac{\left(P(C_k)\prod_{j=1}^{p}[P(x_j|C_k)]\right)/(\mathbf{R})}{\sum_{k=1}^{K}\left(P(C_k)\left(\prod_{j=1}^{p}[P(x_j|C_k)]\right)/\mathbf{R}\right)}$$

Pros and Cons of Naïve Bayes Classifier

- Pros:
  - Simple, efficient and works reasonably well when goal is classification or ranking.
  - Handles large number of categorical inputs very well
  - Have been widely used in Spam filtering, Sentiment Analysis, Text Classification areas
- Cons:
  - Need a lot of records to get good models
  - If a predictor category is not present in the training data, model assumes that a new record with that category of has zero probability!
  - Does not work well if the goal is to get estimates of probability of belonging to a class

Demo Using JMP

- Naïve Bayes is not available in SAS EM but available in SAS Viya
- We will use JMP Pro to demonstrate it
- Dataset: PVA97NK_Imputed