**CS655 – Machine Learning**

**Naïve Bayes and Bayesian Network**

**Week 3 Assignment**

**Student name:** Madeline Lin

Total Points (50 pts)

**1.** **Explain why we cannot avoid altering, even slightly, the character of the data set, when we apply balancing.**

Unbalanced data set refers to those types of data sets where the target class has an uneven distribution of observations, i.e. one class label has a very high number of observations and the other has a very low number of observations. The main problem with imbalanced data set prediction is how accurately are we actually predicting both majority and minority class.

In order to produce an accurate prediction, we are supposed to balance the dataset. Classification decision influenced by preponderance of non-churners in the data set facilitates the need to balance the dataset. When we apply to balance, we try solving the specific business challenges with imbalanced data sets. We cannot avoid altering, the character of the data sets because we need to use balanced sampling methods to reduce the disparity among the proportions of target classes appearing in the training data. Various techniques are used to apply balancing data sets. We usually apply the stratified sampling method, which is 1) partition dataset by target variable, 2) capture all the records containing the minority (target) variable, 3) capture a fixed percentage of the remaining records (the percentage would vary depending on how balanced the final dataset needs to be). After balancing the data, the new dataset will not have exactly the same character as the original data. Also, it is a shame to waste some data. Additionally, the character of the balanced new data will change. For instance, the distribution of variables in the overall balanced dataset will differ (i.e. µ, σ)

**2. Explain the difference in assumptions between naïve Bayes classification and Bayesian networks.**

Naïve Bayes classification assumes attributes are conditionally independent, given the value of the target variable.

P(X1, X2, …, Xd |Yj) = P(X1| Yj) P(X2| Yj)… P(Xd| Yj)

Now we can estimate P(Xi| Yj) for all Xi and Yj combinations from the training data.

New point is classified to Yj if P(Yj) Π P(Xi| Yj) is maximal. (Note the symbol Π means multiplication.)

Naïve Bayes classification is a technique to assign class labels to the samples from the available set of labels. This method assumes each feature’s value as independent and will not consider any correlation or relationship between the features.

Naïve Bayes classification can make decisions with partial information about attributes in the test record. The main issue of the naïve Bayes classification is the performance. We need to find all the probabilities of each and it may take a long time to calculate the probability of a certain event. Also, if any of the probability is zero, then the entire expression becomes zero. We need to use other estimates of conditional probabilities than simple fractions to solve this issue.

Bayesian networks allow joint conditional independencies to be defined among subsets of variables. Bayesian network does not have such assumptions that attributes are conditionally independent. Each variable in a Bayesian network is conditionally independent of its non-descendants in the network, given its parents.



All the dependence in Bayesian Network has to be modeled. The Bayesian network formed can be learned by the machine itself, or can be designed in prior, by the developer, if he has sufficient knowledge of the dependencies. Bayesian networks take the form of a directed acyclic graph (DAG). Directed means that the arcs are traversed in one direction only and acyclic means that no child node cycles back up to any progenitor. Nodes represent variables. Arcs represent the (directed) dependence among the variables. In general, Node A is a parent or immediate predecessor of Node X, and Node X is a descendant of Node A, if there exists a directed arc from A to X.

**3. Explain the difference between the prior and posterior distributions.**

A prior probability represents what is originally believed before new evidence is introduced. A prior probability is the probability that an observation will fall into a group before you collect the data. For example, if you are classifying the buyers of a specific car, you might already know that 60% of purchasers are male and 40% are female.

A posterior probability, in Bayesian statistics, is the revised or updated probability of an event occurring after taking into consideration new information. The posterior probability is calculated by updating the prior probability using Bayes’ theorem. In statistical terms, the posterior probability is the probability of event A occurring given that event B has occurred.

The formula to calculate a posterior probability of A occurring given that B occurred:

P(A∣B) = P(A∩B) / P(B)

= P(A) × P(B∣A) / P(B)

where:

A, B = Events

P(B∣A) = The probability of B occurring given that A is true

P(A) and P(B) = The probabilities of A occurringand B occurring independently of each other​

The posterior probability is thus the resulting distribution, P(A|B). Prior probability represents what is originally believed before new evidence is introduced, and posterior probability takes this new information into account.

Posterior probability distributions should be a better reflection of the underlying truth of a data generating process than the prior probability since the posterior included more information. A posterior probability can subsequently become a prior for a new updated posterior probability as new information arises and is incorporated into the analysis.

**4. Explain in plain English what is meant by the maximum a posteriori classification.**

MAP classification is a probabilistic framework to solve the problem of density estimation. MAP involves calculating a conditional probability of observing the data given a model weighted by a prior probability or belief about the model.

Choose the posterior mode, the value of θ which maximizes *p*(θ |**X**), for an estimate, called *maximum a posteriori* (MAP) method. Bayesian MAPclassification is optimal, which achieves the minimum error rate for all possible classifiers.

MAP classifier may be expressed as below.



**5. In this question, only submit a summary of the article in APA format. You must have short summaries under the headings but do not submit the entire paper.**

**The length of the summary paper is one to two pages. The name of the article is as follow and is in the week 3 folder:**

• **A Bayesian Classification Network-based Learning Status Management System in an Intelligent Classroom**

The topic of this paper is regarding the students’ learning in classroom setting. Overall, the author proposes a learning status management system in an intelligent classroom. The system detects and collects students’ learning status by sensor technology, image recognition technology and a Bayesian classification network.

Learning efficiency is usually influenced by students’ learning status in traditional classrooms. Teachers may not be able to focus on both students’ learning status of each student and their own instructions at the same time. There are several classroom observation instruments which have been designed for human observers to document students’ engagement in class but those are mainly labor dependent, which are time-consuming and laborious. Meanwhile, it is also difficulty for teachers to learn just-in-time results of the observation and change their instructional strategies as the learning status is recognized by observers rather than teachers.

The application of artificial intelligence in education (AIED) in physical classroom settings for enhancing the learning and teaching process has been ongoing. The term *intelligent classroom* refers to a physical classroom that integrates advanced educational technology to improve teachers’ abilities to promote students learning capabilities. To infer students’ learning status from the collected features, a Bayesian classification network was employed. A Bayesian classification network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). A feedback mechanism (LED light) is also included in the system to notify students who had become inattentive. Also, a dashboard is provided to teachers to visualize the real-time learning status of each student so that teachers could then adjust their instructional strategies in a timely fashion to achieve better classroom management.

The Bayesian classification network-based learning status management system consists of sensor technology, image recognition and a feedback mechanism. The four-layer Bayesian classification network is composed of a sensor layer, a feature layer, a behavior layer and a status layer. The sensor layer consists of several types of sensing devices, such as microphone, camera, body temperature. The features of a learner can be captured and recognized via these sensors. The Bayesian classification network proposed here not only uses image recognition technology to incorporate the features that can be recognized from the images/video captured by camera, but considers the information captured from sensors embedded in the classroom and worn by students.

In this study, a learning status management system based on a Bayesian classification network was proposed in an intelligent classroom. Both sensor technology and image recognition technology were employed in the proposed system. Two experiments were conducted to evaluate the accuracy and effectiveness of the proposed system. From the experimental results, the learning status determined by the proposed system was highly correlated to that determined by human observers. Furthermore, the degrees of students’ attention in class could be promoted when the proposed system was enabled. To sum up, the proposed system is helpful to teachers for ensuring more effective classroom management.

**Reference**

Chiu, C.-K., & Tseng, J. C. R. (2021). A Bayesian Classification Network-based Learning Status

Management System in an Intelligent Classroom. *Educational Technology & Society*, *24* (3), 256–267.