# Naїve Bayes

Team: Dylan, Burke, Kristof

Naїve Bayes is a technique used for constructing classifiers founded by Thomas Bayes. The Bayes Theorem formula used by the algorithm is as follows: P(A|B) = P(B|A)P(A)/P(B) with time complexity of O(nM). Some real world applications are sentiment analysis, spam filtering, recommendation systems and more. The classifiers work by correlating use of tokens. Sentiment analysis is an approach to natural language processing (NLP). Spam filtering is used to protect users from unsolicited, unwanted, and virus-infected emails or messages. Recommendation systems include but not limited to generalization of information such as movies sorted into genres and personalizations like filtered content on social media or ads.

The presentation group did very well in articulating their words and made the presentation easy to understand. I do feel when explaining the algorithm before they demonstrated their code, they were moving a little quick;y. Other than the slight hastiness, they did very well in helping us understand naїve bayes algorithm. The real-life examples they used are very intriguing and kept the audience’s attention. Also including the audience in giving them examples to run their algorithm made the presentation fun for the audience to follow along.

# 

# K-Nearest Neighbor Algorithm

Team: Kevin, Jeff, Jacob

K-nearest neighbor is an algorithm that can solve classification and linear regression. Their implementation was used in python. K checks for neighbors and makes decisions. Applications for the algorithm include assessing risk of credit data for banks. The algorithm is easy and simple, adapts well, and has few parameters, so it is very helpful for beginners learning about machine learning algorithms. Some cons include more memory space use, high-dimensionality, and overfitting issues. Their implementation of the algorithm determines whether cancer cells are malignant (cancerous) or benign (not cancerous). The time complexity is O(ndk). A small k value leads to underfitting and poor performance and a large k value can lead to generalized classification which also leads to poor performance.

The group had a lot of good information and were very clear about their points. I left the presentation with a better understanding of k-nearest neighbor algorithm. The real world example was very interesting to see and as a visual learner, I appreciate the data visualizations provided (i.e. the graphs, charts, tables they used to show results).

# Linear Regression

Team: Nicholas, Rafiedul, Luke

Linear regression is used to find the relation between two variables, the correlation between them. Machine learning implementations used are least squares regression which uses a collection of data to find one that better fits than any other line. It is an example of supervised machine learning which is an algorithm trained to find accurate outcomes of data. The algorithm relies on a line of best fit for the most accurate representation of data; achieved by the use of residual plot. The time complexity is O(n2(n+m)) for training and O(n) for prediction where m is the number of columns of the matrix and n is the number of rows.

The performance of the group in their presentation was well-paced and spoke very clearly and loud enough for everyone to hear. The visual representations of the algorithm made it much easier to understand and it was more pleasing to the eye. They simplified a typically practical subject, making the algorithm easy for strangers of the subject to understand.

# Logistic Regression

Team: Alex, Gavin, Dustin

Logistic regression is a statistical model used to estimate probability by identifying classes of data used to predict probability of events. It can be used in binary examples: the medical field, advertisement, and fraudulent transaction. Not to be mistaken with linear regression. Logistic regression focuses on growth that is not constant. There are three types of logistic regression: binary, multinomial, and ordinal. Logistic regression uses the sigmoid function. The implementation was basic but functional, the accuracy needs improvement, needs more visualization and only depends on one variable. The time complexity for training is O(nd), testing is O(dcn), where d is the number of dimensions, c is the number of possible outputs and n is the size of the dataset.

It was very informative and interesting to see the different types of logistic regression and they had a lot of real world examples for each explanation so it was much easier to understand. The visuals provided were helpful and they explained their code very well. I found the comparison of logistic regression vs knn algorithm to be very interesting since it showed how logistic regression could be more accurate than a standard data-driven algorithm.

# ID3 + Machine Learning

Team: Jase & Spencer

Using an iterative dichotomiser 3 (ID3) with Spotify API to analyze information to recommend music and compare it to spotify’s recommendations. ID3 is a greedy algorithm that uses decision trees and was created by Ross Quinlan at the University of Sydney. The time complexity of the algorithm is O(dn2) where d is the number of dimensions and n is the size of the dataset. The ID3 sorts training examples to leaves during training. It uses entropy which is a measurement of randomness and acts as a way for the decision tree to decide when to split the data. The more data, entropy decreases.

The group did a great job articulating their information and they were clear when presenting their points. Carrying the real life example of song recommendation throughout the whole presentation helped with our understanding of ID3. They spoke at a good pace and volume which allowed the audience to better follow along.

# Decision Tree Stumps

Team: Jorge, Will, Vincent

Decision tree stumps are used for binary classification or regression. Classification stumps predict categorical values whereas regression stumps predict numerical values. Decision tree stumps make a decision rule based on a single feature decision. Advantages are robustness, simplicity, and low memory space. The time complexity is O(Nk) where n is the number of samples and k is the number of features.

I felt that the explanation for decision tree stumps and their processes was a little too fast. The presenters could have had more pauses. On the other hand, they spoke clearly and were easy to understand. The real world example for diabetes diagnosis was very interesting and well explained. The code demonstration portion of the presentation was well structured and had nice visuals.

# Apriori Algorithm

Team: Jordan Pinckney

Apriori algorithm is a machine learning algorithm that uses data mining and association rules. It is an iterative approach and puts together item sets from a given data set. The time complexity of finding item sets of size 1 is O(nm). Some uses for it include auto-complete mechanics found in Google, Amazon recommendations, and various stores use it to place items frequently bought closer together when making their mods. There are different sections of the data table: antecedent, consequent, support, confidence, and lift.

Jordan did a great job preventing a project she had to do on her own. She went a little fast in her explanations but still remained clear in them. She was well prepared for the questions that were asked from the audience, so she was clearly knowledgeable on her topic and had put a lot of time and effort into it clearly.

Note on the time complexity: Best case of O(n) and worst case of O(2n)

# Parallel Computing using Thread Model

Team: Brody, Vincent, Meetkumar

Parallel computing uses a multiprocessor to execute programs broken down and executed simultaneously. Two more common computing algorithms are sequential computing and cloud computing. The goal of parallel computing is to increase available computing power to speed up the processes of application processing and problem solving. Four types of parallel computing are bit-level parallelism, instruction-level parallelism, task parallelism, and superword-level parallelism. The group used task parallelism which divides work into smaller tasks which can be executed concurrently and are independent of each other. They used thread model which run concurrently and can communicate with each other and do their individual functions at once.

The group did very well with explaining the concept of parallel computing and was easy to understand as someone that is not entirely familiar with parallel computing as a concept. The slide design was simple yet sophisticated and the provided visuals were helpful to have. I felt their demo was extremely helpful because their chosen language and the use of thread model is a big concept that is gone over in COSC 350 which is very important to know. I definitely found their project useful and interesting.