

TONY NDEREVA MALLIKI
SCR11/0709/2018
SIT407
CAT2

- 1) Discuss the following terms as used in ML
 - a) ML - Machine learning is programming computers to optimize a performance criterion using example data or past experience.
 - b) Noise - Noise is the unwanted data items, features or records which don't help or negatively affect the performance of machine learning algorithms.
 - c) dimensionality reduction - This is a term which refers to techniques that reduce the number of input variables in a dataset.
- 2) Discuss 2 types of Artificial neural networks.
 - (i) feedback ANN - In this ANN, the output returns into the network to accomplish the best-evolved results internally.
 - (ii) Feed-forward ANN - A feed-forward is a basic neural network comprising of an input layer, an output layer and at least one layer of a neuron. Through assessment of its ~~output~~ output by reviewing its input, the intensity of the network can be noticed based on group behaviour of the associated neurons, and the output is decided.
- 3) Describe 5 reasons why NLP is a challenge to researchers.
 - i) Language differences - Unlike numbers and images, language varies from country to country and even within specific regions within the country. As a result, researchers of NLP must work in many languages and without need to undergo retraining each time they encounter a new language.
 - ii) Integration of pre-existing, text-based knowledge - The existing form of knowledge is held in databases, applications, and business processes. As a result, researchers require additional processing to integrate pre-existing information

with the processing of textual content.
iii) extraction of meaning from a variety of complex, multi-format documents - Enterprise information typically resides in PDF documents which are difficult to process. e.g.: These documents come in many formats, leverage structure, to convey meaning, and important information is often included in embedded tables, diagrams, charts and figures.

(iv) Misspellings - Misspellings are an easy challenge for humans to solve, we can quickly link a misspelt word with its correctly spelt equivalent and understand the remainder of the phrase. Misspellings on the other hand, can be more difficult for a machine to detect. Researchers need to employ a natural language processing technology that can identify and progress beyond typical misspelling of terms.

v) Training data - To become proficient in a language, a person must be immersed in it continually for years.

4) Discuss fuzzy logic, how it is useful for commercial and practical purpose. Outline why fuzzy logic is important to ML

- fuzzy logic is a method of reasoning that resembles human reasoning;
- The conventional logic block that a computer can understand takes precise input and produces a definite output as TRUE or false, which is equivalent to human's YES or NO.

- Unlike computers the human decision making includes a range possibility between YES and NO such as:

- ✓ certainly YES
- ✓ Possibly YES
- ✓ Cannot say
- ✓ Possibly NO
- ✓ certainly NO

- Fuzzy logic is useful for commercial and practical purposes.
- (i) It can control machines and consumer products.
 - (ii) It may not give accurate reasoning, but acceptable reasoning.
 - (iii) Fuzzy logic helps to deal with the uncertainty in engineering.

Importance of fuzzy logic in ML are:

- (i) Mathematical concepts within fuzzy reasoning are very simple.
- (ii) Fuzzy logic systems can take imprecise, distorted, noisy input information
- (iii) Fuzzy logic is easy to construct and understand.

5) MUET is looking to hire a person with ML skills to help in development of a system that will be used in the gate to assist identify students as they walk into the campus. As a developer of the system, outline 8 challenges that you are likely to face if you endeavour to develop the system.

- (i) Poor quality of data - One significant issue that the ML profession will face is the absence of good quality data. Unclean and noisy data can make the whole process extremely exhausting.
- (ii) Complexity of Machine learning - The process of Machine learning the program includes transforming, and hence there are high chances of error which makes it complex. It includes analysing the data, removing data bias, training data, applying complex mathematical calculations which these processes might be a challenge for the professional.
- (iii) Imperfections in the algorithm when data grows - Since students continue to enroll, The model used may become inaccurate in the coming future and require further rearrangement.

- v) Slow implementation - The ML models are highly efficient in providing accurate results, but it takes a lot of time. The professional will take a lot of time to use the student's data to train the program.
- vi) Underfitting of training data - The professional may face where data is unable to establish an accurate relationship between input and output variables.
- vii) Overfitting of training data - This refers to a machine learning model trained with a massive amount of data that negatively affect its performance. This means that the algorithm for matching students is trained with noisy and biased data.
- viii) Lack of training data - The school system may not be easily accessible to the professional hence may result to getting data directly from students which may be susceptible to errors and noise.
- * ix) Irrelevant features - feeding bad data that does not make sense to a machine learning model will result in the model outputting an irrelevant output that equally does not make sense.

b) Discuss what is Bayes theorem and explain how it is used in ML

- Bayes theorem is a method to determine conditional probabilities, - that is, the probability of one event occurring given that another event has already occurred.
- Because a conditional probability includes additional conditions, in other words, more data, it contributes to more accurate results.
- Bayes theorem help us to calculate the single term $P(B|A)$ in terms of $P(A|B)$ and $P(A)$. This rule is very helpful in such scenarios

where we have a good probability of $P(A|B)$, $P(B)$ and $P(A)$ and need to determine the forth term.

- Naive Bayes classifier is one of the simplest applications of Bayes theorem which is used in classification algorithms to isolate data as per accuracy, speed and classes.

- Suppose we have vector A with I attributes. It means:

$$A = A_1, A_2, A_3, A_4, \dots, A_i$$

- Further, we have n classes represented as $C_1, C_2, C_3, C_4, \dots, C_n$.

- So with Bayes theorem, we can write it as;

$$P(C_i/A) = [P(A)]$$

$$P(C_i/A) = [P(A/C_i) \times P(C_i)] / P(A)$$

$$P(C_1) = P(C_2) = P(C_3) = P(C_4) = \dots = P(C_n)$$

$$P(A_i/C) = P(A_1/C) \times P(A_2/C) \times P(A_3/C) \dots \times P(A_n/C)$$

- Hence bayes theorem in ML, we can easily describe the possibilities of smaller events

7) MUL limited is having a workshop in which ML approaches are being discussed. As one of the key note speakers in the workshop describe 8 methods of dimensionality reduction.

(i) Principal Component analysis (PCA) - This technique performs a direct mapping of the data to a lesser dimensional space in a way that maximizes the variance of the data in the low-dimensional representation.

(ii) Non-negative matrix factorization (NMF) - This technique breaks down a non-negative matrix into the product of 2 non-negative ones.

(iii) Linear discriminant analysis (LDA) - It is a generalization of Fisher's linear discriminant method that is widely applied in statistics, pattern recognition and ML.

It aims to find a linear combination of features that can characterize or differentiate between 2 or more classes of objects.

iv) Generalized discriminant analysis - It is a nonlinear discriminant analysis that leverages the kernel function operator. Its underlying theory matches very closely to that of support vector machines (SVM), such that the GDA technique helps to map the input vectors into high-dimensional feature space.

v) Missing values Ratio - When you explore a given dataset, you might find that there are some missing values in the dataset. You can then impute the missing values or drop them altogether by using the befitting methods. This approach is best for situations when there are a few missing values.

vi) Low variance filter - When a dataset has constant variables, it is not possible to improve the model's performance because it has 0 variance, so you can set a threshold value to wean out all the constant variables.

vii) High correlation filter - If a dataset consists of data columns having a lot of similar patterns/trends, these data columns are highly likely to contain identical information. For such situations, it's best to use the Pearson's correlation matrix to identify the variables showing a high correlation.

viii) Backward feature elimination - In this technique, you have to begin with all 'n' dimensions. Thus, at a given iteration, you can train a specific classification algorithm is trained on n input features. Now, you have to remove one input feature at a time and train the same model on n-1 input variables n times. Then you remove the input variable whose elimination generates the smallest increase in the error rate, which leaves behind n-1 input feature.

⑧ describe the disadvantages of neural networks.

i) Hardware dependence - Artificial neural networks need processors with parallel processing power, as per their structure.

Therefore, the realization of the equipment is dependent.

ii) Assurance of proper network structure - There is no particular guideline for determining the structure of artificial neural networks. The appropriate network structure is accomplished through experience, trial and error.

iii) difficulty of showing the issue to the network - ANN can work with numerical data. Problems must be converted into numerical values before being introduced to ANN.

The presentation mechanism to be resolved here will directly impact the performance of the network. It relies on the user's abilities.

iv) Unrecognised behaviour of the network - When ANN produces a testing solution, it does not provide insight concerning why and how. It decreases trust in the network.

⑨ With help of a diagram, discuss the fuzzy logic system architecture.

- It has 4 main parts

(i) Fuzzification module - It transforms the system inputs, which are crisp numbers, into fuzzy sets. It splits the input signal into 5 steps such as:

LP - x is large Positive.

MP - x is medium Positive.

S - x is small

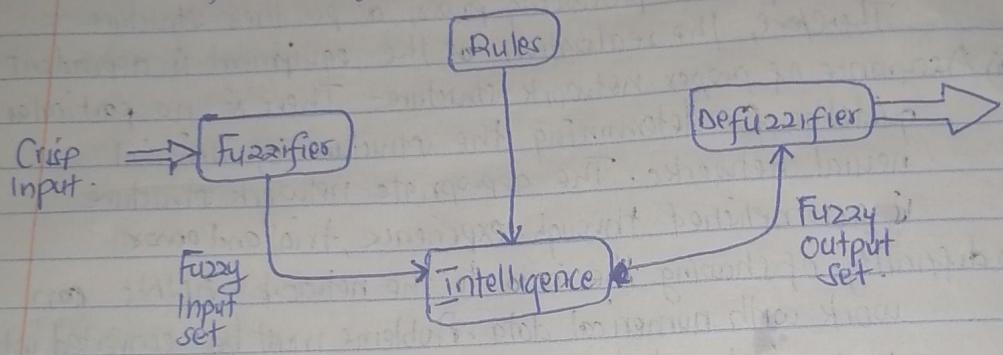
MN - x is medium negative.

LN - x is large Negative.

ii) Knowledge Base - It stores IF-THEN rules provided by experts.

iii) Inference engine - It stimulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.

iv) Defuzzification Module - It transforms the fuzzy set obtained by the inference engine into a crisp value.



(10) Discuss the difference between supervised learning and unsupervised learning algorithms and giving an example in each case.

- A ~~an~~ In supervised learning, a training set of examples with the correct responses is provided and, based on this training set, the algorithm generalizes to respond correctly to all possible inputs.

This is called learning from examples. e.g.: A model is trained that if 10kg of manure produces 100kg of wheat yields; following the same pattern, It will predict that 100kg of manure produces 1000kg of wheat yields.

- In unsupervised learning, correct responses are not provided, but instead the algorithm tries to identify similarities between the inputs so that inputs ~~have~~ that have something in common are categorised together.

(11) Discuss the importance of a target in any machine learning algorithm.

- A target variable of a dataset is the feature of a dataset about which you want to gain a deeper understanding.

- It is important to have a well-defined target since the only thing an algorithm does is learn a function that maps relationships between input data and the target.

- The model's outcome will be meaningless if your target doesn't make sense;

target makes sense -> model makes sense
target doesn't make sense -> model doesn't make sense

(12) Murang'a and company are planning to start a business to supplement their salary which they get from MUR limited.

Describe ways they can use an intelligent agent in the business.

- An intelligent agent is a program that can make decisions or perform service based on its environment

i) for personalised customer experience - These agents are a great way to prevent contextual offers to customers

ii) for streamlined, dynamic governance - Intelligent agents can remove the need for lengthy governance documents by enabling digitised, checkpoints

iii) Better compliance - Intelligent agents keep track of all the decisions they make, which can be used both as a source of continual learning and as a digital audit trail.

Q13 Discuss the following terms:

i) K-NN and K-mean

- K-NN algorithm is a type of supervised ML algorithm which can be used both classification as well as regression predictive problems.

- K-mean algorithm is an unsupervised learning algorithm which groups is used to solve groups the unlabeled dataset into different clusters.

ii) deep learning and shallow learning.

- deep learning is a type of machine learning and artificial intelligence that imitates the way humans gain certain types of knowledge while shallow learning is a method that plateaus at certain level of performance when you add more examples and training data to the network.

iii) Strong AI and weak AI

- Weak AI is a form of AI which focuses on performing a specific task such as answering questions based on user input or playing chess.

- Strong AI is a form of AI which aims to create intelligent machines that are indistinguishable from the human mind.

Q14 Discuss the Conventional neural networks.

- A convolutional Neural network is a deep learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate from one another.
- The pre-processing required in a convNet is much lower compared to other classification algorithms.
- A convNet is able to successfully capture the spatial and temporal dependancies in an image through the application of relevant filters. The architecture performs a better fitting to the

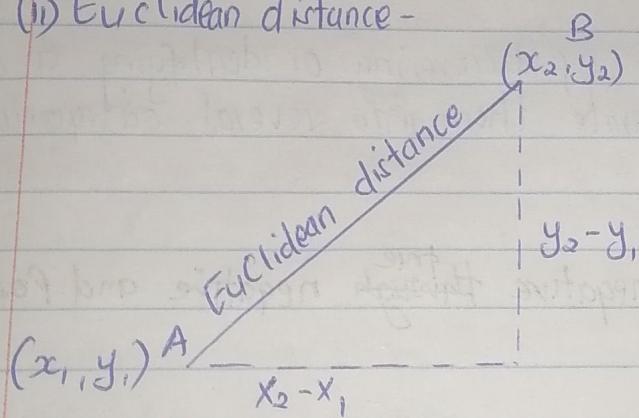
image dataset due to the reduction in the number of parameters involved and reusability of weights

- (15) Discuss the various distance measures used in calculating the closest k-neighbor algorithm.

- i) Manhattan distance - The formula to calculate Manhattan distance is
- $$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

- The left side of the equals sign just means "the distance between point x and point y". The Σ just means "the cumulative sum of each step".

- ii) Euclidean distance -



- The equation at the heart of this distance is the pythagorean theorem.
 $a^2 + b^2 = c^2$.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- iii) Minkowski distance - Is a generalised distance metric across a normed vector space. It can be any function as long as it meets 2 criteria

- (i) the 0 vector (just a vector filled with 0s) will output a length of 0 and,
(ii) every other vector must have a positive length.

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^c \right)^{\frac{1}{c}}$$

- (16) Describe the differences between classification and regression

- Regression are continuous data types and the decision or the outcome variable is continuous.
- It is the method of discovering a function or a model for separating the real values data instead of using distinct values or groups.

- Classification is the process of discovering or identifying a design or role, which helps to separate them into several categorical classes.

- (17) Explain false positive, false negative through true negative and Positive way example.

- A false positive is an outcome where the model incorrectly predicts the positive class eg when pregnancy test is positive but you aren't.
- A false negative is an outcome where the model incorrectly predicts the negative class eg when a person guilty of crimes is acquitted.
- A true positive is an outcome where the model correctly predicts the positive class eg sick people correctly identified as sick.
- A true negative is an outcome where the model correctly predicts the negative class.

(18) Decision
- A
the
on
it
- 1
eo

(19)

⑯ Describe confusion matrix in ML.

- A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. It allows the visualization of an algorithm.
- It allows easy identification of confusion between classes e.g. One class is commonly mislabeled as the other.

⑰ Discuss the difference between KNN and k-mean clustering.

- K-NN is a supervised Machine learning while K-means is an unsupervised machine learning.
- K-NN is a classification or regression machine learning algorithm while K-means is a clustering machine learning algorithm.
- K-NN is a lazy learner while K-means is an eager learner. An eager learner has a model fitting that means a training step but a lazy learner does not have a training phase.
- K-NN performs much better if all of the data have the same scale but this is not true for K-means

⑱ Describe overfitting in ML and how you ensure you are not overfitting in a model.

- This is when our machine learning model tries to cover all the data points or more than the required data points present in the given data.

- b)
- (i) Training with the required amount of data
 - (ii) data simplification
 - (iii) data augmentation
 - (iv) ensembling.

(2D) Describe ensemble learning technique in ML.

- Ensemble learning is a general meta approach to machine learning that seeks better predictive performance by combining the predictions from multiple models.

(2E) Compare and contrast the similarities between bagging and boosting in machine learning.

Similarities:

- Both are the ensemble methods to get N learners from 1 learner
- Both generate several training data sets with random sampling
- Both generate the final result by taking the average of N learners
- Both reduce variance and provide higher scalability.

Differences

- Although they are built independently, but for bagging, boosting tries to add new models which perform well where previous models fail.
- Only boosting tries to reduce bias. Instead, bagging may solve the problem of over-fitting while boosting can increase it.
- Only boosting determine the weight for the data to tip the scales in favor of the most challenging cases.

(6B) Write an SQL query that makes recommendation that your friends liked. Assume you have 2 tables and 2 columns of users and their friends. Other 2 columns of users and pages they liked. It should not recommend pages you already liked.

Table 1 - friends

(user_id, friend_id)

Table 2 - liked

(user_id, page_id)

```
SELECT f.user_id, l.page_id  
FROM friend f JOIN like l  
ON f.friend_id = l.user_id  
WHERE l.page_id NOT IN (SELECT page_id FROM like WHERE  
user_id = f.user_id)
```

④ We have 2 options for serving ads within news feeds:

i) out of every 25 stories 1 will be an advert.

ii) Every story has 4% chance of being an advert.

- for each option either (i) or (ii), what is the expected number of ads shown in 100 news stories?

option 1 = expected number of ads in 100 news stories is equal to 4%

Option 2 = 4% of 100% = $\frac{1}{25}$ = 4 ads.

meaning that one of every 25 stories will be an ad, therefore in 100 newstories, there will be 4 ads.

- Therefore, for each option (i) and (ii), the total number of ads shown in 100 news stories is 4.

⑤ Explain how you will predict who will renew their subscription next month. What data will you need to solve this? What analysis will you do? Will you build a predictive model if so which algorithm?

- Assuming that we are trying to predict renewal rate for spotify subscription, our problem statement is to predict which users will renew their subscription plan next month.

- Data needed, check the number of hours the app is used or is active for each subscriber, the preferences, which songs are played the most, how much time is spent on using the app and how much has the listening rate varied from last month.

- Analysis; for this kind of problem statement, I will use a classification algorithm that classifies the subscribers into two groups
 - users who are likely to subscribe next month.
 - users who are not likely to subscribe next month
- In order to achieve the categorization above, a predictive model that classifies the customers into 2 groups as methods above will be developed or built.
- Algorithms; logistic regression, Random forest, support vector machine, decision trees etc