UNIT 3

1. Explain the term Machine Learning.

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention

~~~~~~

2. What is cost function in linear regression give formula?

cost function — it helps the learner to correct / change behaviour to minimize mistakes. In ML, cost functions are used to estimate how badly models are performing. Put simply, a cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and y. This is typically expressed as a difference or distance between the predicted value and the actual value. The cost function (you may also see this referred to as loss or error.) can be estimated by iteratively running the model to compare estimated predictions against “ground truth” — the known values of y. The objective of a ML model, therefore, is to find parameters, weights or a structure that minimises the cost function.
Example: Cost function(J) of Linear Regression is the Root Mean Squared Error (RMSE) between predicted y value (pred) and true y value (y). Gradient Descent: To update θ1 and θ2 values in order to reduce Cost function (minimizing RMSE value) and achieving the best fit line the model uses Gradient Descent

~~~~~~

3. Mention 3 evaluation metrics formulas in linear regression?

There are 3 main metrics for model evaluation in regression:

1. R Square/Adjusted R Square

2. Mean Square Error(MSE)/Root Mean Square Error(RMSE)

3. Mean Absolute Error(MAE)

R Square/Adjusted R Square: R Square measures how much of variability in dependent variable can be explained by the model. It is square of Correlation Coefficient(R) and that is why it is called R Square.

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

R square formula

R Square is calculated by the sum of squared of prediction error divided by the total sum of square which replace the calculated prediction with mean. R Square value is between 0 to 1 and bigger value indicates a better fit between prediction and actual value. R Square is a good measure to determine how well the model fits the dependent variables. However, it does not take into consideration of overfitting problem

Mean Square Error(MSE)/Root Mean Square Error(RMSE): While R Square is a relative measure of how well the model fits dependent variables, Mean Square Error is an absolute measure of the goodness for the fit.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Mean Absolute Error(MAE): Mean Absolute Error(MAE) is similar to Mean Square Error(MSE). However, instead of the sum of square of error in MSE, MAE is taking the sum of absolute value of error.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Mean Absolute Error formula

Compare to MSE or RMSE, MAE is a more direct representation of sum of error terms. MSE gives larger penalisation to big prediction error by square it while MAE treats all errors the same.

FOR EXTRA POINTS-

Evaluation Metrics in regression models

- **MAE**: mean absolute error $MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$
- **MSE**: mean square error $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- **RMES**: root of mean square error $RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$
- **RAE**: relative average error $RAE = \frac{\sum_{j=1}^{n}|y_j - \hat{y}_j|}{\sum_{j=1}^{n}|y_j - \bar{y}|}$
- **RSE**: relative square error $RSE = \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$
- **R square**： $R^2 = 1 - RSE$

~~~~~~

4. Define Gaussian model?

The Gaussian model is defined by only three parameters: N, μ, and σ, and looks like this: N is the infection rate at its peak, the midpoint of the epidemic. μ is the date of the peak infection rate, and. σ controls the width, the period of time the pandemic is experienced by the country. Gaussian processes are useful in statistical modelling, benefiting from properties inherited from the normal distribution. For example, if a random process is modelled as a Gaussian process, the distributions of various derived quantities can be obtained explicitly.

~~~~~~

5. Define Bernoulli classifier?

The Bernoulli model: There are two different ways we can set up an NB classifier. The model we introduced in the previous section is the multinomial model . It generates one term from the vocabulary in each position of the

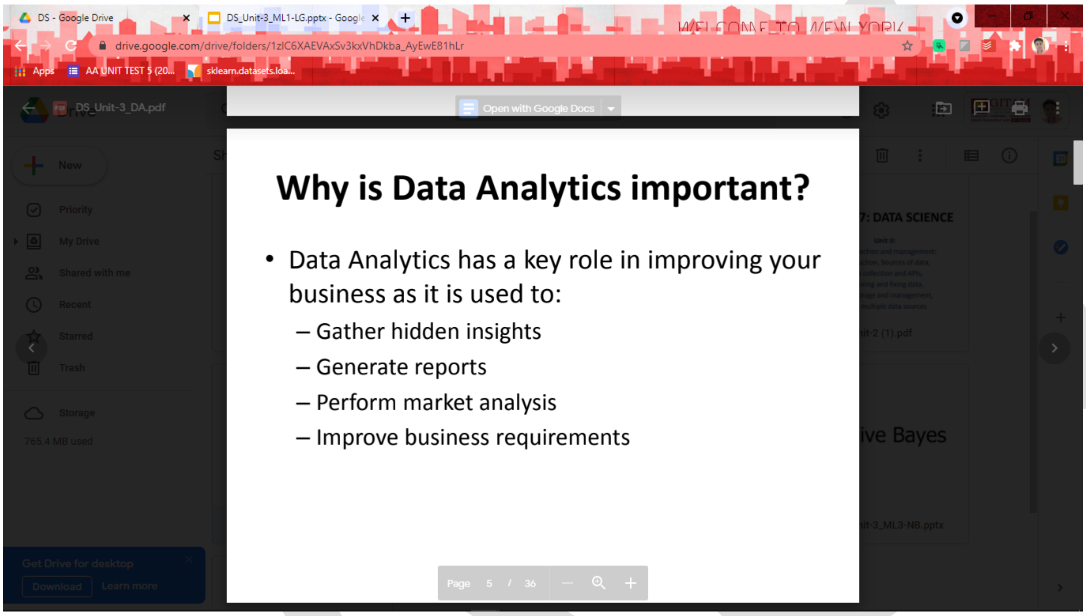
document, where we assume a generative model that will be discussed in more detail in Section 13.4 (see also page 12.1.1).

An alternative to the multinomial model is the multivariate Bernoulli model or Bernoulli model . It is equivalent to the binary independence model of Section 11.3 (page [*]), which generates an indicator for each term of the vocabulary, either 1 indicating presence of the term in the document or 0 indicating absence. Figure 13.3 presents training and testing algorithms for the Bernoulli model. The Bernoulli model has the same time complexity as the multinomial model

Reference: https://nlp.stanford.edu/IR-book/html/htmledition/the-bernoulli-model-1.html

---

6. Why is Data Analytics important?



[Generic thing, elaborate it a little on ur own]

---

7. Define tuning parameters in SVM? [DOUBTFUL]

Support Vector Machine (SVM) is a widely-used supervised machine learning algorithm. It is mostly used in classification tasks but suitable for regression tasks as well. In this post, we dive deep into two important parameters of support vector machines which are C and gamma. So I will assume you have a basic understanding of the algorithm and focus on these parameters. Most of the machine learning and deep learning algorithms have some parameters that can be adjusted which are called hyperparameters. We need to set hyperparameters before we train the models. Hyperparameters are very critical in building robust and accurate models. They help us find the balance between bias and variance and thus, prevent the model from overfitting or underfitting. To be able to adjust the hyperparameters, we need to understand what they mean and how they change a model. It would be a tedious and never-ending task to randomly trying a bunch of hyperparameter values.

Performing the tuning: Now that we have specified a search space and the optimization algorithm, it's time to perform the tuning. We will need to define a resampling strategy and make note of our performance measure. We will use 3-fold cross-validation to assess the quality of a specific parameter setting. For this we need to create a resampling description just like in the resampling part of the tutorial.

~~~~~~

8. Formula for Bayes theorem and mention posterior, prior probability?

Bayes' Theorem is the basic foundation of probability. It is the determination of the conditional probability of an event. This conditional probability is known as a hypothesis. This hypothesis is calculated through previous evidence or knowledge. This conditional probability is the probability of the occurrence of an event, given that some other event has already happened. The formula of Bayes' Theorem involves the **posterior probability P(H | E)** as the product of the probability of hypothesis P(E | H), multiplied by the probability of the hypothesis P(H) and divided by the probability of the evidence P(E).

$$p(H \mid E) = \frac{p(E \mid H)\ p(H)}{p(E)}$$

P(H | E) – This is referred to as **the posterior probability.** Posteriori basically means deriving theory out of given evidence. It denotes the conditional probability of H (hypothesis), given the evidence E.
P(E | H) – This component of our Bayes' Theorem denotes the likelihood. It is the conditional probability of the occurrence of the evidence, given the hypothesis. It calculates the probability of the evidence, considering that the assumed hypothesis holds true.
P(H) – This is referred to as the **prior probability**. It denotes the original probability of the hypothesis H being true before the implementation of Bayes' Theorem. That is, this probability is without the involvement of the data or the evidence.
P(E) – This is the probability of the occurrence of evidence regardless of the hypothesis.

Reference: https://data-flair.training/blogs/bayes-theorem-data-science/#:~:text=The%20formula%20of%20Bayes'%20Theorem,the%20evidence%20P(E).&text=P(H%20%7C%20E)%20%E2%80%93,to%20as%20the%20posterior%20probability

~~~~~~

9. Formula for kernel linear function and polynomial function

1. Linear Kernel: The Linear kernel is the simplest kernel function. It is given by the inner product <x,y> plus an optional constant c. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts, i.e. KPCA with linear kernel is the same as standard PCA. Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set.
k(x, y) = x^T y + c

2. Polynomial Kernel: The Polynomial kernel is a non-stationary kernel. Polynomial kernels are well suited for problems where all the training data is normalized. In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models. Intuitively, the polynomial kernel looks not only at the given features of input samples to determine their similarity, but also combinations of these. In the context of regression analysis, such combinations are known as interaction features. The (implicit) feature space of a polynomial kernel is equivalent to that of polynomial

regression, but without the combinatorial blowup in the number of parameters to be learned. When the input features are binary-valued (booleans), then the features correspond to logical conjunctions of input features

$k(x, y) = (\text{alpha } x^T y + c)^d$

Adjustable parameters are the slope alpha, the constant term c and the polynomial degree d.

~~~~~~

10. Define types of Analytics?

Descriptive analytics: Descriptive analytics answers the question of what happened. Let us bring an example from ScienceSoft's practice: having analyzed monthly revenue and income per product group, and the total quantity of metal parts produced per month, a manufacturer was able to answer a series of 'what happened' questions and decide on focus product categories. Descriptive analytics juggles raw data from multiple data sources to give valuable insights into the past. However, these findings simply signal that something is wrong or right, without explaining why. For this reason, our data consultants don't recommend highly data-driven companies to settle for descriptive analytics only, they'd rather combine it with other types of data analytics.

Diagnostic analytics: At this stage, historical data can be measured against other data to answer the question of why something happened. For example, you can check ScienceSoft's BI demo to see how a retailer can drill the sales and gross profit down to categories to find out why they missed their net profit target. Another flashback to our data analytics projects: in the healthcare industry, customer segmentation coupled with several filters applied (like diagnoses and prescribed medications) allowed identifying the influence of medications. Diagnostic analytics gives in-depth insights into a particular problem. At the same time, a company should have detailed information at their disposal, otherwise, data collection may turn out to be individual for every issue and time-consuming.

Predictive analytics: Predictive analytics tells what is likely to happen. It uses the findings of descriptive and diagnostic analytics to detect clusters and exceptions, and to predict future trends, which makes it a valuable tool for forecasting. Check ScienceSoft's case study to get details on how advanced data analytics allowed a leading FMCG company to predict what they could expect after changing brand positioning. Predictive analytics belongs to advanced analytics types and brings many advantages like sophisticated analysis based on machine or deep learning and proactive approach that predictions enable. However, our data consultants state it clearly: forecasting is just an estimate, the accuracy of which highly depends on data quality and stability of the situation, so it requires careful treatment and continuous optimization.

Prescriptive analytics: The purpose of prescriptive analytics is to literally prescribe what action to take to eliminate a future problem or take full advantage of a promising trend. An example of prescriptive analytics from our project portfolio: a multinational company was able to identify opportunities for repeat purchases based on customer analytics and sales history. Prescriptive analytics uses advanced tools and technologies, like machine learning, business rules and algorithms, which makes it sophisticated to implement and manage. Besides, this state-of-the-art type of data analytics requires not only historical internal data but also external information due to the nature of algorithms it's based on. That is why, before deciding to adopt prescriptive analytics, ScienceSoft strongly recommends weighing the required efforts against an expected added value.

<u>IF 8 MARKS QUESTION, THEN WRITE BELOW ONE ALSO</u>
~~~~~~

Descriptive vs Predictive vs Prescriptive Analytics

Thomas Jefferson said the immortal words – "Not all analytics are created equal."

Descriptive Analytics

Business Intelligence and Data mining

"The simplest class of analytics, one that allows you to condense big data into smaller, more useful nuggets of information."- Dr. Michael Wu

90% of organizations use descriptive analytics.

Analyses the data coming in real-time and historical data for insights on how to approach the future.

Most of the social analytics are descriptive analytics.

Predictive Analytics

Forecasting

"Predictive analytics can only forecast what might happen in the future, because all predictive analytics are probabilistic in nature."- Dr. Michael Wu

Analytics is the next step of data reduction.

Predictive analytics provides answers to questions that cannot be answered by BI.

Predictive analytics can be further categorized as –

What will happen next if <condition>?

Predictive Modelling

Why this actually happened?

Root Cause Analysis

Identifying correlated data

Data Mining

What if the existing trends continue?

Forecasting

What could happen?

Monte-Carlo Simulation

When should an action be invoked to correct a process.

Pattern Indentification and Alerts

Prescriptive Analytics

Simulation and Optimization

Prescriptive analytics is an advanced analytics concept based on –

Optimization that helps achieve the best outcomes.

Stochastic optimization that helps understand how to achieve the best outcome and identify data uncertainties to make better decisions.

Used in producing the credit score which helps financial institutions decide the probability of a customer paying credit bills on time.

Aurora Health Care system saved $6 million annually by using prescriptive analysis to reduce readmission rates by 10%.

DeZyre
www.DeZyre.com

---

11. Explain about Simple Linear Regression model.

Regression models describe the relationship between variables by fitting a line to the observed data. Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line. Regression allows you to estimate how a dependent variable changes as the independent variable(s) change.

Simple linear regression is used to estimate the relationship between two quantitative variables. You can use simple linear regression when you want to know:

- How strong the relationship is between two variables (e.g. the relationship between rainfall and soil erosion).
- The value of the dependent variable at a certain value of the independent variable (e.g. the amount of soil erosion at a certain level of rainfall).

Simple linear regression is a parametric test, meaning that it makes certain assumptions about the data. These **assumptions** are:

- Homogeneity of variance (homoscedasticity): the size of the error in our prediction doesn't change significantly across the values of the independent variable.
- Independence of observations: the observations in the dataset were collected using statistically valid sampling methods, and there are no hidden relationships among observations.
- Normality: The data follows a normal distribution.

The **formula** for a simple linear regression is:

$$y = \beta_0 + \beta_1 X + \varepsilon$$

- y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).
- B0 is the intercept, the predicted value of y when the x is 0.
- B1 is the regression coefficient – how much we expect y to change as x increases.
- x is the independent variable (the variable we expect is influencing y).
- e is the error of the estimate, or how much variation there is in our estimate of the regression coefficient.
- Linear regression finds the line of best fit line through your data by searching for the regression coefficient (B1) that minimizes the total error (e) of the model.

While you can perform a linear regression by hand, this is a tedious process, so most people use statistical programs to help them quickly analyze the data.

---

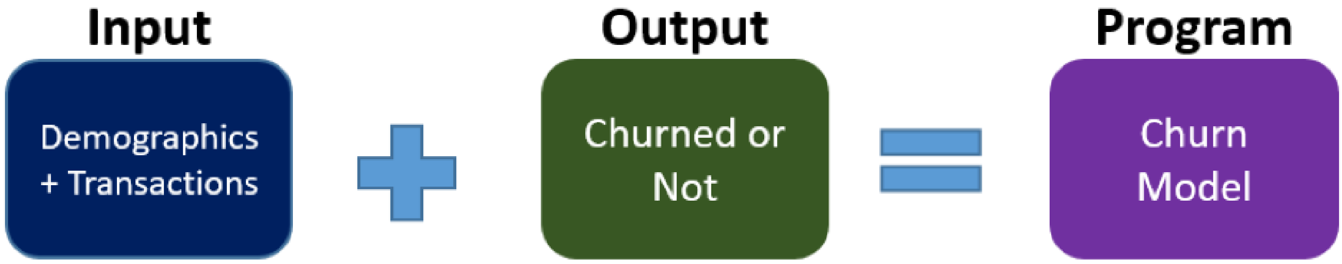
12. Difference between traditional programming and machine learning?

Traditional Programming: Traditional programming is a manual process—meaning a person (programmer) creates the program. But without anyone programming the logic, one has to manually formulate or code rules. Data and program is run on the computer to produce the output.

Machine Learning: Unlike traditional programming, machine learning is an automated process. It can increase the value of your embedded analytics in many areas, including data prep, natural language interfaces, automatic outlier detection, recommendations, and causality and significance detection. All of these features help speed user insights

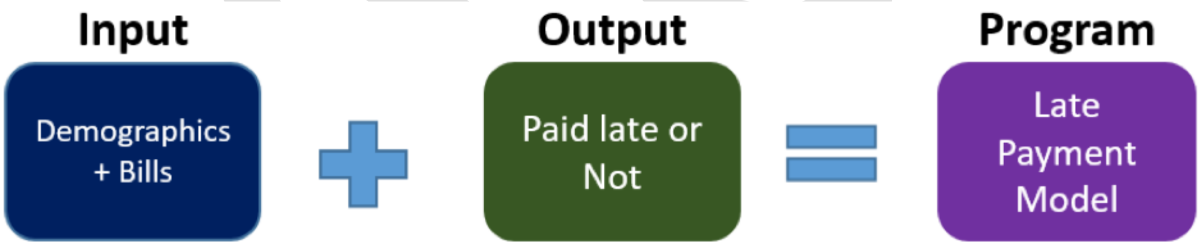
and reduce decision bias. In machine learning, on the other hand, the algorithm automatically formulates the rules from the data.

For example, if you feed in customer demographics and transactions as input data and use historical customer churn rates as your output data, the algorithm will formulate a program that can predict if a customer will churn or not. That program is called a predictive model.



You can use this model to predict business outcomes in any situation where you have input and historical output data:

- Identify the business question you would like to ask.
- Identify the historical input.
- Identify the historically observed output (i.e., data samples for when the condition is true and for when it's false).
- For instance, if you want to predict who will pay the bills late, identify the input (customer demographics, bills) and the output (pay late or not), and let the machine learning use this data to create your model.



Reference: https://www.logianalytics.com/predictive-analytics/machine-learning-vs-traditional-programming/#:~:text=Traditional%20programming%20is%20a%20manual,the%20rules%20from%20the%20data

~~~~~~

13. Explain different types of machine learning?
~~~~~~

	Overview	Process	Subtypes	Examples
Supervised Learning	Majority of algorithms. Machine is trained using **well-labeled data**; inputs and outputs are matched.	Mapping function takes inputs and matches to outputs, creating a target function.	Classification, Regression	Linear regression, Random forest, SVM.
Unsupervised Learning	**Unlabeled data** (inputs only) is analyzed. Learning happens without supervision.	Inputs are used to create a model of the data.	Clustering, Association.	PCA, k-Means, Hierarchical clustering.
Semi supervised	**Some data is labeled,** some not. Goal: better results than labeled data alone. Good for real world data.	Combination of above processes.	All the above.	Self training, Mixture models, Semi-supervised SVM

Supervised	Unsupervised	Semi-Supervised	Reinforcement
• Data has **known labels** or output	• Labels or output unknown • Focus on **finding patterns and gaining insight** from the data	• Labels or output known for **a subset of data** • A blend of supervised and unsupervised learning	• Focus on **making decisions** based on previous experience • Policy-making with feedback
• Insurance underwriting • Fraud detection	• Customer clustering • Association rule mining	• Medical predictions (where tests and expert diagnoses are expensive, and only part of the population receives them)	• Game AI • Complex decision problems • Reward systems

~~~~~~

14. Explain three components of machine learning algorithm?

Every machine learning algorithm has three components:

- Representation: how to represent knowledge. Examples include decision trees, sets of rules, instances, graphical models, neural networks, support vector machines, model ensembles and others.
- Evaluation: the way to evaluate candidate programs (hypotheses). Examples include accuracy, prediction and recall, squared error, likelihood, posterior probability, cost, margin, entropy k-L divergence and others.
- Optimization: the way candidate programs are generated known as the search process. For example combinatorial optimization, convex optimization, constrained optimization.

All machine learning algorithms are combinations of these three components. A framework for understanding all algorithms.

~~~~~~

15. Explain the Data Analytics workflow.

http://www.coordinationtoolkit.org/wp-content/uploads/130907-Data-flow.pdf
Write the topics, and one line description

---

16. What are the different statistical methods that can be applied on data? Explain briefly.

1. Mean: The arithmetic mean, more commonly known as "the average," is the sum of a list of numbers divided by the number of items on the list. The mean is useful in determining the overall trend of a data set or providing a rapid snapshot of your data. Another advantage of the mean is that it's very easy and quick to calculate.

Pitfall: Taken alone, the mean is a dangerous tool. In some data sets, the mean is also closely related to the mode and the median (two other measurements near the average). However, in a data set with a high number of outliers or a skewed distribution, the mean simply doesn't provide the accuracy you need for a nuanced decision.

2. Standard Deviation: The standard deviation, often represented with the Greek letter sigma, is the measure of a spread of data around the mean. A high standard deviation signifies that data is spread more widely from the mean, where a low standard deviation signals that more data align with the mean. In a portfolio of data analysis methods, the standard deviation is useful for quickly determining dispersion of data points.

Pitfall: Just like the mean, the standard deviation is deceptive if taken alone. For example, if the data have a very strange pattern such as a non-normal curve or a large amount of outliers, then the standard deviation won't give you all the information you need.

3. Regression: Regression models the relationships between dependent and explanatory variables, which are usually charted on a scatterplot. The regression line also designates whether those relationships are strong or weak. Regression is commonly taught in high school or college statistics courses with applications for science or business in determining trends over time.

Pitfall: Regression is not very nuanced. Sometimes, the outliers on a scatterplot (and the reasons for them) matter significantly. For example, an outlying data point may represent the input from your most critical supplier or your highest selling product. The nature of a regression line, however, tempts you to ignore these outliers. As an illustration, examine a picture of ANSCOMBE'S QUARTET, in which the data sets have the exact same regression line but include widely different data points.

4. Sample Size Determination: When measuring a large data set or population, like a workforce, you don't always need to collect information from every member of that population – a sample does the job just as well. The trick is to determine the right size for a sample to be accurate. Using proportion and standard deviation methods, you are able to accurately determine the right sample size you need to make your data collection statistically significant.

Pitfall: When studying a new, untested variable in a population, your proportion equations might need to rely on certain assumptions. However, these assumptions might be completely inaccurate. This error is then passed along to your sample size determination and then onto the rest of your statistical data analysis

5. Hypothesis Testing: Also commonly called t testing, hypothesis testing assesses if a certain premise is actually true for your data set or population. In data analysis and statistics, you consider the result of a hypothesis test statistically significant if the results couldn't have happened by random chance. Hypothesis tests are used in everything from science and research to business and economic

Pitfall: To be rigorous, hypothesis tests need to watch out for common errors. For example, the placebo effect occurs when participants falsely expect a certain result and then perceive (or actually attain) that result. Another common error is the Hawthorne effect (or observer effect), which happens when participants skew results because they know they are being studied.

---

17. Explain about central limit theorem.

The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement text annotation indicator, then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually n > 30). If the population is normal, then the theorem holds true even for samples smaller than 30. In fact, this also holds true even if the population is binomial, provided that min(np, n(1-p))> 5, where n is the sample size and p is the probability of success in the population. This means that we can use the normal probability model to quantify uncertainty when making inferences about a population mean based on the sample mean.

$$\mu_{\bar{X}} = \mu$$

For the random samples we take from the population, we can compute the mean of the sample means:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Before illustrating the use of the Central Limit Theorem (CLT) we will first illustrate the result. In order for the result of the CLT to hold, the sample must be sufficiently large (n > 30). Again, there are two exceptions to this. If the population is normal, then the result holds for samples of any size (i..e, the sampling distribution of the sample means will be approximately normal even for samples of size less than 30).

Reference [if asked for 8 marks then]: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability12.html#:~:text=The%20central%20limit%20theorem%20states,will%20be%20approximately%20normally%20distributed

~~~~~~

18. Define Bayes law. How does Naïve Bayes solve spam filter problem?

Bayes' theorem allows you to update predicted probabilities of an event by incorporating new information. Bayes' theorem was named after 18th-century mathematician Thomas Bayes. It is often employed in finance in updating risk evaluation.

Bayes' theorem relies on incorporating prior probability distributions in order to generate posterior probabilities. Prior probability, in Bayesian statistical inference, is the probability of an event before new data is collected. This is the best rational assessment of the probability of an outcome based on the current knowledge before an experiment is performed. Posterior probability is the revised probability of an event occurring after taking into consideration new information. Posterior probability is calculated by updating the prior probability by using Bayes' theorem. In statistical terms, the posterior probability is the probability of event A occurring given that event B has occurred.
~~~~~~

Formula For Bayes' Theorem

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}$$

where:

$P(A) =$ The probability of A occurring

$P(B) =$ The probability of B occurring

$P(A|B) =$ The probability of A given B

$P(B|A) =$ The probability of B given A

$P\left(A \bigcap B\right)) =$ The probability of both A and B occurring

Reference: https://www.investopedia.com/terms/b/bayes-theorem.asp

The concept of spam filtering is simple - detect spam emails from authentic (non-spam/ham) emails. To do this, the goal would be to get a measure of how 'spammy' an incoming email is. The extended form of Bayes' Rule comes into play here. With Bayes' Rule, we want to find the probability an email is spam, given it contains certain words. We do this by finding the probability that each word in the email is spam, and then multiply these probabilities together to get the overall email spam metric to be used in classification. The the probability of an email being spam S given a certain word W appears is defined by the left hand side of the above equation Pr(S|W).

The right hand side gives the formula to compute this probability. This is:

the probability the word occurs in the email given it is a spam email Pr(W|S) multiplied by the probability of an email being spam Pr(S),

divided the probability the word occurs in the email given it is a spam email multiplied by the probability of an email being spam,

plus the probability the word occurs in the email given it is a non-spam email Pr(W|¬S) multiplied by the probability of an email being non-spam Pr(¬S).

Probabilities can range between 0 and 1. For this spam filter, we will define that any email with a total 'spaminess' metric of over 0.5 (50%) will be deemed a spam email.

When the Pr(S|W) has been found for each word in the email, they are multiplied together to give the overall probability that the email is spam. If this probability is over the 'spam threshold' of 0.5, the email is classified as a spam email.

Reference : https://towardsdatascience.com/na%C3%AFve-bayes-spam-filter-from-scratch-12970ad3dae7

[OPEN IN INCOGNITO]

~~~~~~

19. Explain Linear regression and Discuss the advantages and disadvantages.

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models are target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Please refer Linear Regression for complete reference. Linear Regression is a great tool to analyze the relationships among the variables but it isn't recommended for most
~~~~~~

practical applications because it over-simplifies real-world problems by assuming a linear relationship among the variables.

Advantages

Linear Regression is simple to implement and easier to interpret the output coefficients.

When you know the relationship between the independent and dependent variable have a linear relationship, this algorithm is the best to use because of it's less complexity to compared to other algorithms.

Linear Regression is susceptible to over-fitting but it can be avoided using some dimensionality reduction techniques, regularization (L1 and L2) techniques and cross-validation.

Disadvantages

On the other hand in linear regression technique outliers can have huge effects on the regression and boundaries are linear in this technique.

Diversely, linear regression assumes a linear relationship between dependent and independent variables. That means it assumes that there is a straight-line relationship between them. It assumes independence between attributes.

But then linear regression also looks at a relationship between the mean of the dependent variables and the independent variables. Just as the mean is not a complete description of a single variable, linear regression is not a complete description of relationships among variables.

REFER Qn 11 ALSO [IF TIME]

Reference : https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/

~~~~~~

20. Explain SVM and Discuss the advantages and disadvantages.

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text.

ALGORITHM : https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

Types of SVM

SVM can be of two types:

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Advantages:

- SVM works relatively well when there is a clear margin of separation between classes.
- SVM is more effective in high dimensional spaces.
- SVM is effective in cases where the number of dimensions is greater than the number of samples.
- SVM is relatively memory efficient

Disadvantages:

- SVM algorithm is not suitable for large data sets.
~~~~~~

- SVM does not perform very well when the data set has more noise i.e. target classes are overlapping.
- In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.
- As the support vector classifier works by putting data points, above and below the classifying hyperplane there is no probabilistic explanation for the classification.

SVM algorithm can be used for Face detection, image classification, text categorization, etc

~~~~~~

21. Explain Naive Bayes and Discuss the advantages and disadvantages.

QN 8 +

SVM Advantages

- SVM’s are very good when we have no idea on the data.
- Works well with even unstructured and semi structured data like text, Images and trees.
- The kernel trick is real strength of SVM. With an appropriate kernel function, we can solve any complex problem.
- Unlike in neural networks, SVM is not solved for local optima.
- It scales relatively well to high dimensional data.
- SVM models have generalization in practice, the risk of over-fitting is less in SVM.
- SVM is always compared with ANN. When compared to ANN models, SVMs give better results.

SVM Disadvantages

- Choosing a “good” kernel function is not easy.
- Long training time for large datasets.
- Difficult to understand and interpret the final model, variable weights and individual impact.
- Since the final model is not so easy to see, we can not do small calibrations to the model hence its tough to incorporate our business logic.
- The SVM hyper parameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact

SVM Application

- Protein Structure Prediction
- Intrusion Detection
- Handwriting Recognition
- Detecting Steganography in digital images
- Breast Cancer Diagnosis
- Almost all the applications where ANN is used

~~~~~~