

Question 6
Which step / phase of the learning workflow that defines the potential accuracy of your project?

Data Acquisition

Question 7
In which phase of the learning workflow one needs to break dataset? and what are these sets?

Data Preprocessing (and Feature Engineering)

0.25 points Save Answer
What does mse stand for?

Mean Square Error

How we can prove that a given ML model will work well on new data and not only on the training data?

We need to check the accuracy of the model against the test data set. This proves if the model works well on new data or if it only works on the training data.

Why do we need to shuffle the data before starting the training process?

So, the ML algorithm doesn't pick up a pattern that isn't really there! Because the data in certain order.

The set of parameters that are selected based on their influence on the model are called

hyperparameters

2. Differentiate between classification and regression with an example. (1.5 Marks)

Classification vs. Regression:

Classification	Regression
<ul style="list-style-type: none"> Classification is the task of predicting a discrete class label In a classification problem data is labelled into one of two or more classes A classification problem with two classes is called binary, more than two classes is called a multi-class classification Classifying an email as spam or non-spam is an example of a classification problem 	<ul style="list-style-type: none"> Regression is the task of predicting a continuous quantity A regression problem requires the prediction of a quantity A regression problem with multiple input variables is called a multivariate regression problem Predicting the price of a stock over a period of time is a regression problem

QUESTION 3
Identify which ML technique adapts a policy for mapping from states actions that tells you what to do in a given state

- Supervised Learning
- Semi supervised learning
- Reinforcement learning
- UnSupervised Learning

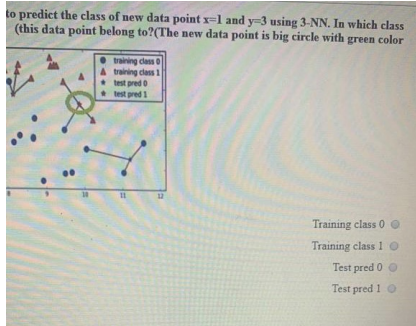
QUESTION 6
Differentiate between univariate, model-based selection and iterative feature selection.

Consider adult incomes in the United States dataset derived from the 1994 census database, which consists of continuous features- age and hours-per-week and categorical features- workclass, education, sex, and occupation

Analyze, which of the below technique is used to replace categorical variables with one or more new features of continuous variables

Univariate Nonlinear Transformations
One-hot encoding
Binning
ANOVA

- السؤال 10
- 1 درجات
1. which feature selection technique builds a series of models with varying numbers of features
- Non-Univariate Transformation
 - Iterative Selection
 - Univariate Statistics
 - Model-based Selection



QUESTION 6
Consider a binary classification problem to predict whether a customer has more than 3 children or is not divorced, then they want to buy a boat". Which of the following is true?

- The rule generalises to new data
- The rule is able to capture all the aspects of and variability in the data
- The Rule seems too simple and underfitting occurs
- The Rule seems too complex and overfitting occurs

السؤال 12

Python uses _____ to indicate blocks

- Newline
- semicolon
- Square brackets
- indentation

السؤال 13

Machine learning technique learns a function $f(x)$ to predict y given x

- Supervised Learning
- Probabilistic modelling
- Reinforcement learning
- UnSupervised Learning

QUESTION 12

Given a IRIS flower dataset which contains class labels machine learning problems it belongs

- Binary and Multiclass classification
- Regression
- Binary Classification
- Multiclass Classification

QUESTION 13

Discuss the pros and cons of K-NN

Let's say, you are working with the synthetic dataset of counts, which of the below technique can be applied to overcome the challenges in Univariate nonlinear transformation

both B and C



Whether a customer will buy a boat. Given " If the customer is older than 45, and divorced, then they want to buy a boat". Which of the below claims are correct

- The Rule seems too complex and overfitting occurs
- The Rule seems too simple and underfitting occurs
- The rule generalises to new data
- The rule is able to capture all the aspects of and variability in the data

Parasitic gaps occur on average in 1/100,000 sentences. Maggie Louise Gal (aka "ML" Gal) has developed a machine learning approach to identify parasitic gaps. If a sentence has a parasitic gap, it correctly identifies it 97% of the time. If it doesn't it will incorrectly say it does with probability 0.003. Suppose we run it on a sentence and the algorithm says it is a parasitic gap, what is the probability it actually is " with probability 0.003.

- 0.003 -
- 0.004 -
- 0.005 -
- 0.003 -

السؤال 12

dataset which contains class labels {Iris Satosa, Iris Virginia, Iris Versicolor}. Identify which of the following supervised machine learning problems it belongs

السؤال 9 1 درجات حفظ الإجابة

In scikit-learn, which function given below converts categorical variables to dummy continuous variables

- pd.get_dummies
- Both A and Cs
- OneHotEncoder
- data_dummies.head



F1 represents color of mobile phones for certain brand. Recognize, which of

- Feature F1 is an example of discrete variable
- Feature F1 is an example of categorial variable
- Feature F1 is an example of continuous variables
- Feature F1 is an example of discrete and continuous variables

QUESTION 1

Identify which machine learning model can handle both categorical and numerical

- A. Logistic Regression
- B. Neural Networks
- C. Decision trees
- D. Naive Bayes



Given the below table

ML Pass AND Eng Pass	Eng Pass (Eng Pass)
true, true	48
true, false	41
false, true	44
false, false	47

What is: $P(ML\ Pass = true | Eng\ Pass = true)$?

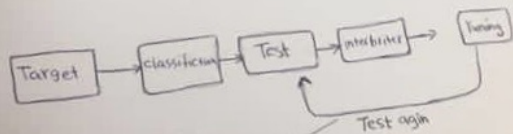
- A. 0.125
- B. 0.215
- C. 0.59
- D. 0.95

10. Parasitic gaps occur on average in 1/100,000 sentences. Maggie Louise Gal (aka "ML" Gal) has developed a machine learning approach to identify parasitic gaps. If a sentence has a parasitic gap, it correctly identifies it 97% of the time. If it doesn't, it will incorrectly say it does with probability 0.003. Suppose we run it on a sentence and the algorithm says it is a parasitic gap, what is the probability it actually is?

- A. -0.003
- B. -0.002
- C. -0.004
- D. -0.005

PART II: (10 Marks)

1. Discuss with a neat figure the sequence of steps in Machine learning workflow. (1.5 Marks)



Quiz:

- If you're given some scenario to specify whether, it's Supervised or unsupervised?

Scenario 1: Facebook recognizes your friend in a picture from an album of tagged photographs

Explanation:

It is supervised learning. Here Facebook is using tagged photos to recognize the person. Therefore, the tagged photos become the labels of the pictures and we know that when the machine is learning from labeled data, it is supervised learning.

Quiz:

- If you're given some scenario to specify whether it's Supervised or unsupervised?

Scenario 2: Recommending new songs based on someone's past music choices

Explanation:

It is supervised learning. The model is training a classifier on pre-existing labels (genres of songs). This is what Netflix, Pandora, and Spotify do all the time, they collect the songs/movies that you like already, evaluate the features based on your likes/dislikes and then recommend new movies/songs based on similar features.

Quiz:

- If you're given some scenario to specify whether it's Supervised or unsupervised?

Scenario 3: Analyze bank data for suspicious-looking transactions and flag the fraud transactions

Explanation:

It is unsupervised learning. In this case, the suspicious transactions are not defined, hence there are no labels of "fraud" and "not fraud". The model tries to identify outliers by looking at anomalous transactions and flags them as 'fraud'.

Exercise 1:

Suppose we have a linear regression problem, perform two iterations using gradient descent algorithm. Show the final values of θ_0 and θ_1 , consider initial values for

$$\begin{aligned} \theta_0 &= 2 \\ \theta_1 &= 3 \\ \alpha &= 0.5 \end{aligned}$$

And consider the dataset below:

	x_1	y
1	12	5
2	5	2

$m=2$

Gradient Descent

Previously ($n=1$):

Repeat {
 $\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$
 $\frac{\partial}{\partial \theta_0} J(\theta)$

$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x^{(i)}$
(simultaneously update θ_0, θ_1)

New algorithm ($n \geq 1$):

Repeat {
 $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$
(simultaneously update θ_j for $j = 0, \dots, n$)

Handwritten solution for Exercise 1:

Round 1:

$\theta_0 = 2$
 $\theta_1 = 3$
 $\alpha = 0.5$

1) $h(x_1) = \theta_0 + \theta_1 x_1$
 $h(x_0) = 2 + 3 \cdot 12 = 38$
 $h(x_1) = 2 + 3 \cdot 5 = 17$

2) $h(x^{(i)} - y^{(i)})$
 $i=1 \Rightarrow 38 - 5 = 33$
 $i=2 \Rightarrow 17 - 2 = 15$

3) $\theta_0 := \theta_0 - \alpha \frac{1}{m} \cdot ((h(x_0) - y_0) + (h(x_1) - y_1))$
 $= 2 - \frac{0.5}{2} (33 + 15) = -10$

$\theta_1 := \theta_1 - \alpha \frac{1}{m} \cdot ((h(x_0) - y_0) \cdot x_0 + (h(x_1) - y_1) \cdot x_1)$
 $= 3 - \frac{0.5}{2} (33 \cdot 12 + 15 \cdot 5)$
 $= 3 - \frac{0.5}{2} (396 + 75)$
 $= -114.75$

Round 2:

$\theta_0 = 484.4$
 $\theta_1 = 4793.43$