42577 - Introduction to Business Analytics
Quiz

Name and Surname:	Student Number:

Instructions: The quiz consists of 12 multiple-choice questions and 1 open-text question. The multiple-choice questions may have multiple options that are correct. The questions that have multiple correct options are indicated in the text. Please circle all the correct options. You receive 1 point for a correct answer and 0 points for a wrong answer (in case multiple options are correct you need to select all the correct options to get 1 point). There is no negative marking. Please indicate your name, surname, and student number.

1. The code below shows two ways of dealing with categorical data. Elaborate on the difference and discuss which one would you choose if you wanted to include this variable in a logistic regression model.

```python
import pandas as pd

data = {'Category': ['A', 'B', 'A', 'C', 'B', 'C']}
df = pd.DataFrame(data)

# Create dummy variables
dummy_df_option_1 = pd.get_dummies(df['Category'], prefix='Category')
dummy_df_option_2 = pd.get_dummies(df['Category'], drop_first=True, prefix='Category')

# Concatenate the dummy variables with the original DataFrame
df1 = pd.concat([df, dummy_df_option_1], axis=1)
df2 = pd.concat([df, dummy_df_option_2], axis=1)
```

```
In [27]: print(df1)
  Category  Category_A  Category_B  Category_C
0        A           1           0           0
1        B           0           1           0
2        A           1           0           0
3        C           0           0           1
4        B           0           1           0
5        C           0           0           1
```

```
In [28]: print(df2)
  Category  Category_B  Category_C
0        A           0           0
1        B           1           0
2        A           0           0
3        C           0           1
4        B           1           0
5        C           0           1
```

<u>Answer</u>:

2. You have a dataset with a single variable X1 and a response variable Y. You find that at each value of X1 the distribution of Y is log-normal. You wish to use linear regression to model Y as a function of X1. What would you do?
   A) Fit Y as a linear function of X1.
   B) Fit exp(Y) as a linear function of X1.
   C) Linear regression is a bad idea since the OLS assumptions are violated.
   D) Fit log(Y) as a linear function of X1.

3. Assume that you are given a data set containing 1000 observations and 750 features (Xs) along with a single response variable Y. You estimate 3 OLS linear regression models to predict Y as a function of the

features (with no regularization): M1 contains 5 features, M2 contains 20 features, and M3 contains 750 features. You find that the test and the training MSEs are as follows:

Model	Training MSE	Test MSE
M1 (5 features)	150	170
M2 (20 features)	40	45
M3 (750 features)	38	200

Which of the following do you think is most likely:

A) M1 has a high bias but low variance, M3 has low bias but high variance, M2 balances bias and variance is the preferred model out of the three.
B) M2 has a high bias but low variance, M1 has low bias but high variance, M3 balances bias and variance is the preferred model out of the three.
C) M3 has a high bias but low variance, M2 has low bias but high variance, M1 balances bias and variance is the preferred model out of the three.
D) We cannot make any statement about which is the preferable model based on the information provided in the table.

4. In logistic regression for a binary target variable, which of the following statement(s) are true (**more than one option can be correct**):

A) The odds-ratio is modeled as a linear function of the features.
B) The log odds-ratio is modeled as a linear function of the features.
C) The probability of success (y=1) is modeled as a linear function of the features.
D) We can use both categorical and continuous variables as features.

5. You have a data set with a binary response variable Y and a single feature X. You estimate a logistic regression model and find that the coefficient of X in the model is +2. This means that (**more than one option can be correct)**:

A) Increasing X by one unit increases the log-odds ratio by 2 units.
B) Increasing X by one unit increases the log-odds ratio by 1 unit.
C) Decreasing X by one unit decreases the log-odds ratio by 2 units.
D) Increasing X by one unit decreases the log-odds ratio by 2 units.

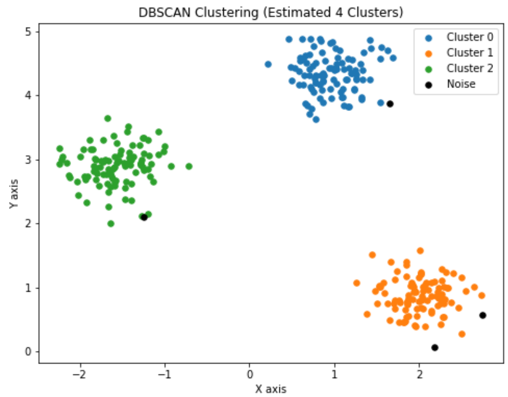
6. You are interested in modeling a classification problem with a binary response variable (Y = 1 is a success or a positive, Y=0 is a failure or a negative). Assume that the cost of predicting a success (Y=1) when the true value is a failure (Y=0) is very large (higher than that of predicting a failure when the true value is success). You train a logistic regression model and a SVM and obtain the following metrics:

Model	Precision	Recall
Logistic Regression (LR)	0.95	0.7
SVM	0.75	0.75

Which of the models would you recommend using based on the above information?

A) I would recommend LR.
B) I would prefer SVM.
C) Both are equally preferable.
D) Both are equally bad and should not be used.

7. Look at the figure below. It corresponds to the results obtained using DBSCAN on a bi-dimensional database. Imagine that you want to perform the same operation using k-means (with k=3) instead of DBSCAN to identify Cluster 0, 1, and 2. Which of the following statements is correct?



A) K-means (with k=3) can recreate the exact same clusters as in the figure.
B) K-means (with k=3) can recreate approximately the same clusters as in the figure.
C) Due to the strict assumption on the data, K-means should not be used for this type of data.

8. What is the fundamental difference between supervised and unsupervised machine learning?

A) In supervised learning, the algorithm makes predictions without any training data, while in unsupervised learning, the algorithm uses labeled data for training.
B) In supervised learning, the algorithm learns patterns and relationships in data without the need for a target variable, while in unsupervised learning, the algorithm requires a target variable for training.
C) In supervised learning, the algorithm is provided with input-output pairs during training, allowing it to learn to make predictions, while in unsupervised learning, the algorithm works with unlabeled data and tries to discover inherent patterns or structures.
D) Supervised and unsupervised learning are interchangeable terms that refer to the same machine learning paradigm.

9. In Principal Component Analysis (PCA) with eigenvalue decomposition, what does the first eigenvalue of the covariance matrix represent?

A) The total variance in the data.
B) The variance explained by the first principal component.
C) The sum of squared differences between data points and their means.
D) The correlation between the features in the dataset.

10. What is the vanishing gradient problem in deep neural networks, and which activation function was introduced to mitigate it?

A) It's a problem of overly large gradients during training, and the Rectified Linear Unit (ReLU) was introduced to mitigate it.
B) It's a problem of gradients becoming too small, making weight updates negligible, and the ReLU activation function was introduced to mitigate it.
C) It's a problem of overfitting, and the Dropout regularization technique was introduced to mitigate it.
D) It's a problem of exploding gradients during training, and the Tanh activation function was introduced to mitigate it.

11. What is the primary purpose of dropout in neural networks?

A) To increase the depth of the network for better representation.
B) To introduce noise into the training process to improve model generalization.

C) To decrease the learning rate during training for faster convergence.
D) To reduce the number of parameters in the model for computational efficiency.

12. When training a neural network using gradient-based optimization methods, which statement is true regarding local optima?

    A) Local optima are not a concern in neural networks because they always converge to the global optimum.
    B) Neural networks are highly likely to get stuck in local optima, making it difficult to find the best model.
    C) Neural networks are guaranteed to reach the global optimum in most cases.
    D) Local optima only exist in shallow neural networks, not in deep ones.
