Week 1 Evaluation - 1.5 hours - pass/fail

NAME: ___Jeong ganghyeon_____

Week 1 Evaluation

Question 1 - Visualization, Data Pre-processing, Feature Selection

a) **Which of the following plots is good at showing the outliers? (i)**
   i) Box Plot
   ii) Pie Chart
   iii) Line Chart
   iv) Heatmaps

b) **What of the following is not an objective of data cleaning? (i)**
   i) Creating new data records.
   ii) Removing duplicate records.
   iii) Handling missing data.
   iv) Correcting inaccurate values

c) **Why do we use feature selection in machine learning? (iii)**
   i) To change data into a machine learning format.
   ii) To pick the best machine learning method.
   iii) To choose the important features.
   iv) To fill in missing data.

d) **Coding Question:**
   Consider the following DataFrame

	Name	Age	Score
0	Alice	25	90
1	Bob	18	82
2	Charlie	12	70
3	David	30	95

```
# Use 'loc' to select the row with the name 'Charlie' and all columns.
df.loc[df[“Name”] == [“Charlie”], :]
```

Question 2 - Model Evaluation, Hyperparameter Tuning, Bagging and Boosting

a) **You've trained a deep decision tree on your dataset and noticed that it performs really well on the training data but poorly on the validation data. (ii)**
Which ensemble technique might help remedy this, bagging or boosting?
   i) Bagging
   ii) Boosting

b) **Which technique is commonly used for finding the best hyperparameter tuning. (ii)**
   i) Feature Enginnering
   ii) Cross-Validation
   iii) Data Preprocessing
   iv) Model Inference

c) **Which of the following is not typically considered as an evaluation metrics for classification models? (iii)**
   i) Precision
   ii) Recall
   iii) Feature Importance
   iv) F1-Score

d) **Coding Question:**
Complete the blanks below to enable the code to run successfully.

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.svm import SVC

# Load the Iris dataset
data = load_iris()
X, y = data.data, data.target

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Initialize the SVM classifier
svm_classifier = SVC()

# Define the hyperparameter grid for GridSearchCV
param_grid = {
    'kernel': ['linear', 'rbf', 'poly', 'sigmoid'],
    'C': [0.1, 1, 10, 100],
    'gamma': [1, 0.1, 0.01, 0.001]
}

# Create the GridSearchCV object with cross-validation (e.g., 5-fold cross-
```

PDF-XChange
Click to buy NOW!
www.tracker-software.com

```
validation)
grid_search = GridSearchCV(estimator = svm.classifier, param_grid = param_grid,
cv=5)

# Perform the hyperparameter tuning on the training data
grid_search.fit(X_train, y_train)

# Get the best hyperparameters from the GridSearchCV
best_params = grid_search.best_params_
print("Best hyperparameters:", best_params)
```

Question 3 - Supervised learning (linear regression)

a) **What are the two common types of supervised learning? (choose 2) (i, iii)**
   i) Classification
   ii) Clustering
   iii) Regression

b) **Which of the following statements is true about supervised learning? (iv)**
   i) It is a type of machine learning where the model is trained on labeled data.
   ii) It is a type of machine learning where the model does not require any training data.
   iii) It is a type of machine learning where a human must intervene during the training process.
   iv) It is a type of machine learning that does not require labeled data.

c) **For linear regression, which of the following are the inputs (also known as features) that are fed into the model and which the model makes a prediction on? (i)**

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = X\beta + \varepsilon$$

   i) x
   ii) β
   iii) y
   iv) ε

d) **Coding question:**

Please complete code by replacing all `### EXERCISE CODE HERE` with your code.

```
"""
Your task is to complete the code to predict the weight of someone using a
height and gender.
"""

import pandas as pd
import numpy as np
# weight-height dataset
dataset =
pd.read_csv("https://gist.githubusercontent.com/nstokoe/7d4717e96c21b8ad04ec91f3
61b000cb/raw/bf95a2e30fceb9f2ae990eac8379fc7d844a0196/weight-height.csv")
gender_mapping = {'Male': 0, 'Female': 1}
dataset["Gender"] = dataset["Gender"].map(gender_mapping)
display(dataset)

#put the columns of ones into the front of the dataset
dataset.insert(0, "ones", 1)
```

PDF-XChange
Click to buy NOW!
www.tracker-software.com

```
X = dataset[["ones", "Gender", "Height"]].values
y = dataset["Weight"].values

# Calculating the coefficients using the normal equation: (X^T * X)^(-1) * X^T *
y
coefficients = np.linalg.inv(X.T.dot(X)).dot(X.T).dot(y)

print(coefficients)

"""predict the weight of someone with height of 70 and who have gender `male`"""

X_test = np.array([1,0,70])

# Making predictions for the test set
y_prediction = X_test.dot(coefficients)
print(y_prediction)
```

Please test your code here:
https://github.com/MLcmore2023/MLcmore2023/blob/main/day3_am_morning/question.ipynb

Question 4 - Supervised learning (logistic regression)

a) **What is a gradient? (ii)**
   i) The rate of change of the model's accuracy during training.
   ii) A vector that points in the direction of maximum increase of a function (usually the loss function)
   iii) A type of data structure used to store multiple values of the same data type for model training
   iv) A measure of the angle between two intersecting vectors (usually the training dataset and the testing dataset)

b) **What is the purpose of gradient descent? (iii)**
   i) Visualizing data patterns
   ii) Generating random numbers to initialize a model
   iii) Minimizing the loss function to improve the model
   iv) Converting categorical data to numerical data

c) **When training the logistic regression model, we use gradient descent to minimize the loss function. What happens if the loss function became zero? (iii)**
   i) This means the model's prediction will always be zero.
   ii) This means our code is wrong, because it is impossible.
   iii) This means the model now fit the dataset really well.
   iv) This means the model now fit the test dataset really well.

d) **The output of logistic regression is: (ii)**
   i) A continuous value
   ii) A probability score
   iii) A categorical label
   iv) A binary value

e) **Coding Question**

```
#apply logistic regression to the following dataset, to bank notes as real or
fake.
#The Banknote Authentication Dataset is a collection of data used for binary
classification. It contains features extracted from images of genuine and
counterfeit banknotes, captured using various sensors.

import pandas as pd
dataset =
pd.read_csv("https://raw.githubusercontent.com/Dataweekends/zero_to_deep_learnin
g_video/master/data/banknotes.csv")
display(dataset)

X = dataset.drop(columns=['class'])  # Replace 'target_column_name' with the
actual target column name
y = dataset['class']
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
X_train, X_test, y_train, y_test = X_train.to_numpy(), X_test.to_numpy(),
y_train.to_numpy(), y_test.to_numpy()

logistic_reg = LogisticRegression()
logistic_reg.fit(X_train, y_train)

y_pred = logistic_reg.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

Please test your code here:
https://github.com/MLcmore2023/MLcmore2023/blob/main/day3_am_morning/logistic-regression-demo.ipynb

Question 5 - Supervised learning (decision trees / random forest)

a) **How is the splitting attribute chosen in a decision tree? (ii)**
   i) Randomly selecting an attribute from the dataset.
   ii) Choosing the attribute that has the highest information gain or the lowest Gini impurity.
   iii) Using gradient descent to find the best attribute.
   iv) Selecting the attribute that has the highest correlation with the target variable.

b) **What is the purpose of a cost function? (ii)**
   i) The cost function is used to calculate the total number of data points in the training set.
   ii) The cost function is a measure of how good or bad the decision tree fits the training data.
   iii) The cost function is used to evaluate the complexity of the decision tree.
   iv) The cost function is only applicable to unsupervised learning algorithms, not decision trees.

c) **What is ensemble learning? (iii)**
   i) A single machine learning model
   ii) A group of unrelated models
   iii) A technique that combines predictions from multiple models
   iv) A method to visualize data

d) **In Random Forest, what is bootstrapping used for? (iii)**
   i) Enhancing model interpretability
   ii) Preventing overfitting and speed up training
   iii) Generating different random subsets of data for training
   iv) Improving the convergence of the algorithm

e) **Coding Question**

```
#Make a predict_1_sample function, which takes in 1 sonar data x and returns the prediction
def predict_1_sample(x, random_forest_model):
  tree_predictions = []
  for tree in random_forest_model.estimators_:
    tree_predictions.append(tree.predict([x])[0])
 final_prediction = sum(tree_predictions) / len(tree_predictions)
 return final_prediction


# Make predictions on the test set
predictions = []
for x in X_test:
  p = predict(x, random_forest_model)
  predictions.append(p)
print(predictions)
```

Please test your code here:
https://github.com/MLcmore2023/MLcmore2023/blob/main/day3_pm_afternoon/random-forest-demo.ipynb

Question 6 - Deep learning (multi layer preceptron networks)

A. **Why are activation functions important in MLPs? (c)**
   a. They allow for the introduction of linear transformations.
   b. They determine the number of layers in the network.
   c. They introduce non-linearity, enabling the model to learn complex patterns.
   d. They reduce the computational complexity of the network.

B. **Which of the following is NOT a property of the sigmoid function? (c)**
   a. Differentiable for all real numbers
   b. Output range between 0 and 1
   c. The result of *sigmoid*(2000) is twice as large as *sigmoid*(1000)
   d. Continuous for all real numbers

C. **Which of the following best describes a Multi-layer Perceptron (MLP)? (d)**
   a. A single-layer feedforward neural network.
   b. A neural network that primarily uses convolutional layers.
   c. A type of recurrent neural network optimized for time-series data.
   d. A feedforward neural network with one or more hidden layers.

D. **Coding question**

Feedforward

$$a' = \sigma(wa + b)$$

a is the vector of activations of the n-th layer of neurons. To obtain a' (n+1 th layer), we multiply a by the weight matrix w, and add the vector b of biases. We then apply the function σ elementwise to every entry in the vector wa+b.

```
# write a feedforward function that works on neural networks of 5 layers.
def feedforward(a, biases, weights):
    #layer 1 --> 2
    b = biases[0]
    w = weights[0]
    a_layer2 = sigmoid(np.dot(w, a) + b)

    # EXERCISE code
```

Question 7 - Deep learning (CNN)

a) **Which of the following is a primary reason for using convolutional layers in a neural network when processing image data? (iii)**
   i) They increase the number of parameters in the network.
   ii) They allow the network to be fully connected.
   iii) They are designed to detect local patterns, like edges and textures.
   iv) They ensure that the network only processes black and white images.

b) **In a CNN, what does the term 'pooling' generally refer to? (iii)**
   i) Increasing the depth of the network by adding more convolutional layers.
   ii) The process of periodically saving the network's weights to avoid data loss.
   iii) Reducing the spatial dimensions of the data by taking the average or maximum value from a group of neighboring pixels.
   iv) Applying a filter to increase the sharpness of the image before processing.

c) **Which of the following is NOT a common activation function used in CNNs? (iv)**
   i) ReLU (Rectified Linear Unit)
   ii) Tanh
   iii) Leaky ReLU
   iv) Linear Regression

d) **In the context of CNNs, what does a 'stride' refer to? (iv)**
   i) The depth of the convolutional filter.
   ii) The size of the pooling window.
   iii) The number of layers in the network.
   iv) The number of pixels the convolutional filter moves after processing a section of the input.

e) **Coding question:**

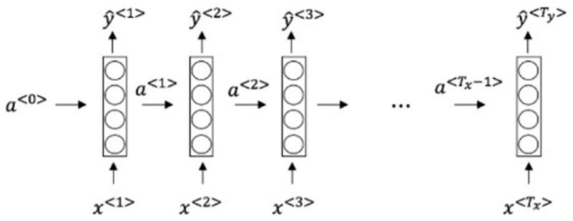
You are building a simple CNN using TensorFlow's Keras API to classify images. Fill in the blank to add a convolutional layer to the model:

```
import tensorflow as tf

model = tf.keras.Sequential()

# Add a convolutional layer with 32 filters, a kernel size of (3,3), and 'relu'
activation:
model.add(tf.keras.layers.Conv2D(32, (3, 3), activation='relu'))
```

Question 8 - Deep learning (RNN)

a) **In a Recurrent Neural Network, what is the primary purpose of the hidden layer? (ii)**
   i) To process the input data and produce the final output.
   ii) To store sequential information for memory
   iii) To apply activation functions to the input features.
   iv) To directly connect with the output layer.

b) **Suppose you have the following RNN architecture. In what cases would you use it? (i)**



   i) Action Recognition in a video
   ii) Sentiment Classification
   iii) Trigger word Detection
   iv) Named Entity Recognition

c) **What is the main advantage of using a Recurrent Neural Network (RNN) compared to a traditional feedforward neural network? (i)**
   i) RNNs can predict on input sequences of any length
   ii) RNNs have more layers, leading to better performance.
   iii) RNNs are faster in training due to parallelization.
   iv) RNNs require less memory for computation.

d) **To which of the following you would use many-to-one architecture of RNN? (i)**
   i) Sentiment Classification
   ii) Question Answering
   iii) Named Entity Recognition
   iv) Music Generation

e) **You are training an RNN based architecture and realized during the training that the weights have a value of NaN. Which of these is a possible reason for the same? (iv)**
   i) Because the RNN cell uses sigmoid activation function.
   ii) Because it uses ReLU activation function.
   iii) Vanishing gradient problem
   iv) Exploding gradient problem

f) **Coding question:**

```
# Feed forward manual example
i_weight = np.random.rand(1,10)
h_weight = ### EXERCISE CODE HERE
h_bias = ### EXERCISE CODE HERE
o_weight = np.random.rand(10,1)
o_bias = np.random.rand(1,10)
```

PDF-XChange
Click to buy NOW!
www.tracker-software.com

```
temps = [66, 70, 62, 74, 80,68, 67]
#predict the temperature for tomorrow (day 8)
### EXERCISE CODE HERE
```