#### MANE 4962 HW 3

Lucas Zhou 662005044

## File Loaded

```
from google.colab import files
uploaded = files.upload()

Browse... 2 files selected.
housing_prices.txt(text/plain) - 1408 bytes, last modified: n/a - 100% done
HW3.1.jpg(image/jpeg) - 7952 bytes, last modified: n/a - 100% done
```

Saving housing\_prices.txt to housing\_prices.txt

Saving HW3.1.jpg to HW3.1.jpg

## Problem 1:

Image segmentation is a process to highlight useful regions in images. Use the skimage in module to load the following image. Afterwards, segment the image into multiple useful regions using the k-means clustering method. The segmented image should highlight, for example, the dashboard, the driver's arms, cars ahead etc., by grouping similar pixels together. You do not need to split the data into train and test set for this problem.

```
from skimage.io import imread
from skimage.color import rgb2lab
from sklearn.cluster import KMeans
import numpy as np
import matplotlib.pyplot as plt

# Load the image
image = imread('/content/HW3.1.jpg')
# Convert the image from RGB to Lab color space
```

```
image_lab = rgb2lab(image)

# Reshape the image to a 2D array of Lab color values
pixels = image_lab.reshape((-1, 3))

# Apply K-means clustering
kmeans = KMeans(n_clusters=6, random_state=42)
kmeans.fit(pixels)
labels = kmeans.labels_

# Reshape labels back to the original image dimensions
segmented_image = labels.reshape(image.shape[0], image.shape[1])

# Display the segmented image
plt.imshow(segmented_image, cmap='viridis')
plt.axis('off')
plt.show()
```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change warnings.warn(



#### v Problem 2.

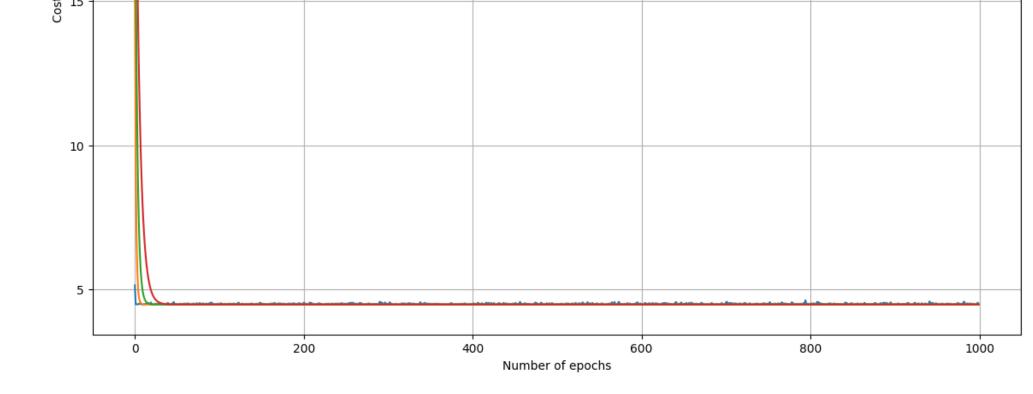
i iobiciti Z.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
# Load the data
data path = '/content/housing prices.txt'
data = pd.read_csv(data_path, header=None, delimiter=',', skiprows=1)
# Extracting features (population) and target values (price)
X = data.iloc[:, 0].values # Population
y = data.iloc[:, 1].values # Housing price
# Normalize the features (population)
X_{normalized} = (X - np.mean(X)) / np.std(X)
# Adding a column of ones to the features to act as bias (intercept term)
X_b = np.c_[np.ones((len(X), 1)), X_normalized]
# Parameters for mini-batch gradient descent
learning rate = 0.01
n = 1000
batch_sizes = [1, 5, 10, 20]
cost history batch = {}
# Mini-batch gradient descent function
def mini_batch_gradient_descent(X, y, learning_rate, batch_size, n_epochs):
    m = len(y)
    theta = np.random.randn(2, 1) # Random initialization of weights
    cost_history = []
    for epoch in range(n_epochs):
        shuffled_indices = np.random.permutation(m)
        X_shuffled = X[shuffled_indices]
        y_shuffled = y[shuffled_indices].reshape(-1, 1)
        for i in range(0, m, batch_size):
            xi = X_shuffled[i:i + batch_size]
            yi = y shuffled[i:i + batch size]
            gradients = 2/batch size * xi.T.dot(xi.dot(theta) - yi)
```

```
theta = theta - learning_rate * gradients
            y_predict = X_b.dot(theta)
            cost = (1/(2*m)) * np.sum(np.square(y_predict - y.reshape(-1, 1)))
        cost_history.append(cost)
    return theta, cost_history
# Plotting the cost function for each batch size
plt.figure(figsize=(14, 10))
for batch_size in batch_sizes:
    theta, cost_history = mini_batch_gradient_descent(X_b, y, learning_rate, batch_size, n_epochs)
    cost_history_batch[batch_size] = cost_history
    plt.plot(range(len(cost_history)), cost_history, label=f'Batch size {batch_size}')
plt.xlabel('Number of epochs')
plt.ylabel('Cost J')
plt.title('Cost J vs. Number of epochs for different batch sizes')
plt.legend()
plt.grid()
plt.show()
```







# With extra loop

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Load the data
data_path = '/content/housing_prices.txt'
data = pd.read_csv(data_path, header=None, delimiter=',', skiprows=1)

# Extracting features (population) and target values (price)
X = data.iloc[:, 0].values # Population
y = data.iloc[:, 1].values # Housing price
```

```
# Normalize the features (population)
X_{normalized} = (X - np.mean(X)) / np.std(X)
# Adding a column of ones to the features to act as bias (intercept term)
X_b = np.c_[np.ones((len(X), 1)), X_normalized]
# Parameters for mini-batch gradient descent
learning_rate = 0.01
n = 1000
batch_sizes = [1, 5, 10, 20]
cost history_batch = {}
# Mini-batch gradient descent function
def mini_batch_gradient_descent_with_extra_loop(X, y, learning_rate, batch_size, n_epochs, extra_loops):
    m = len(y)
    theta = np.random.randn(2, 1) # Random initialization of weights
    cost_history = []
    for epoch in range(n_epochs):
        shuffled_indices = np.random.permutation(m)
        X_shuffled = X[shuffled_indices]
        y_shuffled = y[shuffled_indices].reshape(-1, 1)
        for i in range(0, m, batch_size):
            xi = X_shuffled[i:i + batch_size]
            yi = y_shuffled[i:i + batch_size]
            for _ in range(extra_loops): # Extra loop
                gradients = 2/batch size * xi.T.dot(xi.dot(theta) - yi)
                theta = theta - learning rate * gradients
            y_predict = X.dot(theta)
            cost = (1/(2*m)) * np.sum(np.square(y_predict - y.reshape(-1, 1)))
        cost_history.append(cost)
    return theta, cost_history
# Example usage
learning_rate = 0.01
n_{epochs} = 1000
batch_sizes = [1, 5, 10, 20]
extra loops = 5 # Number of times the undate sten is repeated per batch
```

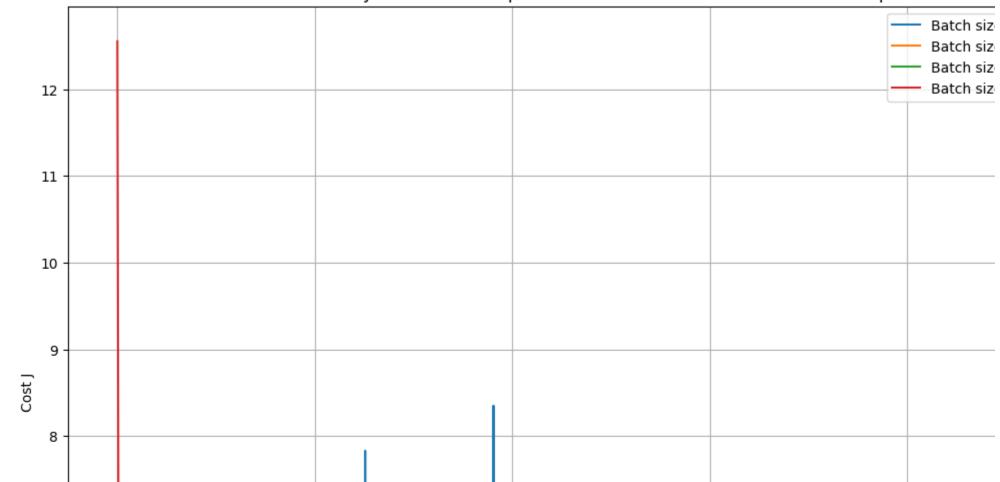
```
cost_history_batch = {}

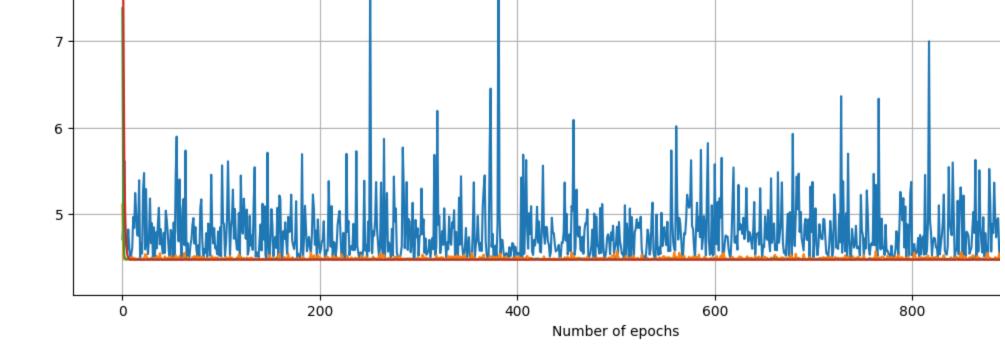
plt.figure(figsize=(14, 10))

for batch_size in batch_sizes:
    theta, cost_history = mini_batch_gradient_descent_with_extra_loop(X_b, y, learning_rate, batch_size, n_epochs, extra_loops)
    cost_history_batch[batch_size] = cost_history
    plt.plot(range(len(cost_history)), cost_history, label=f'Batch size {batch_size} with extra loops')

plt.xlabel('Number of epochs')
plt.ylabel('Cost J')
plt.title('Cost J vs. Number of epochs for different batch sizes with extra loops')
plt.legend()
plt.grid()
plt.show()
```







## Problem 3:

```
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import RFE
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import numpy as np

# Load dataset
data = load_breast_cancer()

# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(data.data, data.target, test_size=0.3, random_state=42) # 70% training and 30%

# Create a logistic regression model
logreg = LogisticRegression()

# Recursive Feature Elimination for feature selection
rfe = RFE(logreg, n_features_to_select=2)
rfa = rfe fit(Y train y train)
```

```
# Identify which two features are considered to be the most important.
selected features = np.where(rfe.support == True)[0]
print("Selected features:", selected features)
# Retrain the model with the selected features
logreg.fit(X train[:, selected features], y train)
# Predict test set
y pred = logreg.predict(X test[:, selected features])
# Evaluate the model
print("Confusion Matrix:\n", confusion matrix(y test, y pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
print("Accuracy Score:", accuracy score(y test, y pred))
     /usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
     STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
     Increase the number of iterations (max iter) or scale the data as shown in:
         https://scikit-learn.org/stable/modules/preprocessing.html
     Please also refer to the documentation for alternative solver options:
         https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
       n_iter_i = _check_optimize_result(
     /usr/local/lib/python3.10/dist-packages/sklearn/linear model/ logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
     STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
     Increase the number of iterations (max iter) or scale the data as shown in:
         https://scikit-learn.org/stable/modules/preprocessing.html
     Please also refer to the documentation for alternative solver options:
         https://scikit-learn.org/stable/modules/linear model.html#logistic-regression
       n_iter_i = _check_optimize result(
     /usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
     STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
     Increase the number of iterations (max iter) or scale the data as shown in:
         https://scikit-learn.org/stable/modules/preprocessing.html
     Please also refer to the documentation for alternative solver options:
         https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
       n_iter_i = _check_optimize_result(
     /usr/local/lib/python3.10/dist-packages/sklearn/linear model/ logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
     STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
   https://scikit-learn.org/stable/modules/linear model.html#logistic-regression
 n iter i = check optimize result(
/usr/local/lib/python3.10/dist-packages/sklearn/linear model/ logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
Increase the number of iterations (max iter) or scale the data as shown in:
   https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear model.html#logistic-regression
 n iter i = check optimize result(
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
Increase the number of iterations (max iter) or scale the data as shown in:
   https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
   https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
 n_iter_i = _check_optimize result(
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
Increase the number of iterations (max_iter) or scale the data as shown in:
   https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
   https://scikit-learn.org/stable/modules/linear model.html#logistic-regression
 n iter i = check optimize result(
/usr/local/lib/python3.10/dist-packages/sklearn/linear model/ logistic.py:458: ConvergenceWarning: lbfgs failed to converge (sta
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

### Problem 4:

Double-click (or enter) to edit

```
import tensorflow as tf
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import pandas as pd
```

```
# Load and prepare the dataset
data = pd.read csv('/content/housing prices.txt')
X = data.iloc[:, 0].values.reshape(-1, 1) # Features
y = data.iloc[:, 1].values.reshape(-1, 1) # Labels
# Normalize the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
# Split the dataset into a training set and a validation set
X train, X val, y train, y val = train test split(X scaled, y, test size=0.3, random state=42)
# Build the neural network
model = tf.keras.Sequential([
    tf.keras.layers.Dense(2, activation='relu', input shape=(X train.shape[1],)),
    tf.keras.layers.Dense(1)
])
# Compile the network
model.compile(optimizer='sgd', loss='mse')
# Train the network
history = model.fit(X train, y train, epochs=100, validation data=(X val, y val))
# Plot the training and validation loss
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val loss'], label='Validation Loss')
plt.legend()
plt.show()
# Predict the price of a house in a city with population of 165,000 (scaled accordingly)
population 165k scaled = scaler.transform([[16.5]]) # The population feature scaled
predicted price = model.predict(population 165k scaled)
print("Predicted price of house:", predicted price)
# Calculate a regression metric (Root Mean Squared Error, for example)
y_pred = model.predict(X_val)
mse = tf.keras.losses.MeanSquaredError()
rmse = np.sqrt(mse(y_val, y_pred).numpy())
print("Root Mean Squared Error on Validation Set:", rmse)
```

import numpy as np

```
Epoch 1/100
Epoch 2/100
Epoch 3/100
Epoch 4/100
Epoch 5/100
Epoch 6/100
Epoch 7/100
Epoch 8/100
Epoch 9/100
Epoch 10/100
Epoch 11/100
Epoch 12/100
Epoch 13/100
Epoch 14/100
Epoch 15/100
Epoch 16/100
Epoch 17/100
Epoch 18/100
Epoch 19/100
Epoch 20/100
Epoch 21/100
Epoch 22/100
Epoch 23/100
```

```
Epoch 24/100
Epoch 25/100
Epoch 26/100
Epoch 27/100
Epoch 28/100
Epoch 29/100
Epoch 30/100
Epoch 31/100
Epoch 32/100
Epoch 33/100
Epoch 34/100
Epoch 35/100
Epoch 36/100
Epoch 37/100
Epoch 38/100
Epoch 39/100
Epoch 40/100
Epoch 41/100
Epoch 42/100
Epoch 43/100
Epoch 44/100
Epoch 45/100
Epoch 46/100
```

```
Epoch 47/100
Epoch 48/100
Epoch 49/100
Epoch 50/100
Epoch 51/100
Epoch 52/100
Epoch 53/100
3/3 [=========================== ] - 0s 129ms/step - loss: 8.1181 - val loss: 11.0247
Epoch 54/100
Epoch 55/100
Epoch 56/100
Epoch 57/100
Epoch 58/100
Epoch 59/100
Epoch 60/100
Epoch 61/100
Epoch 62/100
Epoch 63/100
Epoch 64/100
Epoch 65/100
Epoch 66/100
Epoch 67/100
Epoch 68/100
Epoch 69/100
```

```
Epoch 70/100
Epoch 71/100
Epoch 72/100
Epoch 73/100
Epoch 74/100
Epoch 75/100
Epoch 76/100
Epoch 77/100
Epoch 78/100
Epoch 79/100
Epoch 80/100
Epoch 81/100
Epoch 82/100
Epoch 83/100
Epoch 84/100
Epoch 85/100
Epoch 86/100
Epoch 87/100
Epoch 88/100
Epoch 89/100
Epoch 90/100
Epoch 91/100
Epoch 92/100
```

```
Epoch 93/100
Epoch 94/100
Epoch 95/100
Epoch 96/100
3/3 [=============== ] - 0s 14ms/step - loss: 8.0860 - val loss: 11.1255
Epoch 97/100
Epoch 98/100
Epoch 99/100
Epoch 100/100
Training Loss
60
              Validation Loss
50
40
30
20
10
```

60

80

100

1/1 [======] - 0s 83ms/step Predicted price of house: [[16.034851]] 1/1 [======] - 0s 21ms/step

40

20

0

Root Mean Squared Error on Validation Set: 3.400657