ASSIGNMENT 2

Q1. SVM Regression (3 Points)

You are given a dataset that contains information about the prices of used cars based on their age and mileage. Your goal is to build an SVM regression model to predict the prices of cars. You need to perform the following tasks:

1. Data Exploration:

- Load and explore the dataset.
- Visualize the data to understand the relationship between age, mileage, and car prices.

2. Data Preprocessing:

- Split the dataset into training and testing sets.
- Standardize the features (age and mileage) to have zero mean and unit variance.

3. SVM Regression:

- Train an SVM regression model using a linear kernel.
- Tune the hyperparameters (e.g., C, epsilon) using cross-validation or grid search to find the best model.

4. Model Evaluation:

- Evaluate the model's performance on the testing set using appropriate regression metrics (e.g., Mean Absolute Error, Mean Squared Error, R-squared).
- Visualize the model's predictions against the actual car prices.

5. Discussion:

- Discuss the results and the impact of hyperparameter tuning on the model's performance.
- Compare the SVM regression model with other regression techniques (e.g., linear regression, decision tree regression) in terms of predictive accuracy.

6. Conclusion:

- Summarize your findings and provide insights into the key factors that influence used car prices based on your SVM regression model.

Additional Information:

The dataset is provided in a CSV file containing columns for "Age," "Mileage," and "Price." You can use Python and relevant libraries (e.g., scikit-learn) for data analysis, modelling, and visualization. Make sure to provide code, plots, and explanations for each step in your assignment. Please ensure to submit a well-documented report that includes code, visualizations, and a detailed explanation of your approach and findings.

Q2. Implement SVM Classifier (3 Points)

Implement an SVM classifier to classify images from the CIFAR-10 dataset into their respective classes.

Instructions:

- **Dataset Loading:** Load the CIFAR-10 dataset. You can access this dataset using popular deep learning libraries like PyTorch or TensorFlow, or download it from the CIFAR-10 website.

- **Data Preprocessing:** Preprocess the dataset by normalizing the images and flattening them into feature vectors. **(Extra Credit:** Augment additional images to the dataset using flipping and rotation). Split the dataset into training and testing sets in the ratio 80:20 using stratified sampling.

- **Feature Extraction:** Implement feature extraction techniques if needed. For eg: perform Principal Component Analysis (PCA) for dimensionality reduction.

- **SVM Classifier Implementation:**

  1. Implement a linear SVM classifier using a library like Scikit-Learn. Train the classifier on the training dataset. Tune the hyperparameters of the classifier by changing the value of C = [1,10, 100].

  2. Implement an SVM classifier with RBF kernel. Train the classifier on the training dataset. Tune the hyperparameters of the classifier by changing the values of $\gamma = [10^{-3}, 10^{-4}]$, C= [1,10,100].

- **Model Evaluation:** Evaluate the performance of both SVM classifier on the test dataset using classification metrics such as accuracy, precision, recall, and F1-score.

- **Visualization:** Visualize some example images that were correctly classified and some that were misclassified to understand the model's performance.

- **Write a Report:** Create a report based on your findings.

Q3. Support Vector Regression (SVR) -based stock price prediction with the influence of news events in Python (4 Points)

Instructions:

1. **Obtain API Key:**
   - Go to the [News API website](#).
   - Sign up for an account to obtain your API key.
   - Replace "your_news_api_key_here" in the code with your actual News API key.

2. **Download Historical Stock Data:**
   - Use the yfinance library to download historical stock price data for a chosen company. For this assignment, we will focus on Apple Inc. (AAPL) stock.
   - Set the start and end dates for data retrieval (e.g., start_date = "2022-10-31", end_date = "2023-10-31").
3. **Download News Data:**
   - Utilize the NewsApiClient from the newsapi-python library to retrieve news articles related to the chosen company.
   - Specify the company name and date range for news data retrieval.
4. **Data Merging:**
   - Merge the stock data and news data based on the publication date.
5. **Feature Engineering:**
   - Create a new feature based on news sentiment or other relevant information extracted from the news articles.
6. **Data Splitting:**
   - Split the dataset into training and testing sets (e.g., 80% training, 20% testing).
7. **Model Building:**
   - Build an SVR model (Support Vector Regression) **from scratch** with the 'rbf' kernel.
8. **Model Training and Evaluation:**
   - Train the SVR model on the training data and make predictions of Adjusted closing price and Opening price of next 30 days.
   - Evaluate the model's performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2).
   - Build a pickle file of the model and run this for three different companies.
9. **Assignment Submission:**
   - Submit the Python code along with any relevant comments or explanations of the code. Also make a report on the same.

DEADLINE FOR SUBMISSION IS 13 NOVEMBER 11.59 PM.