



Task Documentation

Ahmed Shehata
UmedMi

Overview

In this documentation i will explain the 3 notebooks required in the task
I will explain the steps, difficulties, and how I overcame it.

Goals

Document the 3 projects in the task

Specifications

We have 3 projects:

- U Intern Task
- Digit-recognizer
- Fake-and-real-news Detection

First one: U Intern Task

In this notebook, we have some small required methods:-

First Task is to build a method to compute the value of $a+aa+aaa+aaaa$ with a given digit as the value of a

So, I have used the %s operator which is most often used in conjunction with the % operator for strings.

Second Task is to write a program to print the running time of execution of "1+1" for 100 times.

so , I have used the timeit python library to compute the start and the end of the execution.

Third Task is to write a program to sort the (name, age, height) tuples by ascending order

So, I have used a for loop to loop over the string and split it by ' , ' then store it in my tuple and use the sorted method to sort in ascending order.

Fourth Task is to Write a Python program to count the number of characters (character frequency) in a string.

So, have used the dictionary data structure in python, to use the key for looping over the string and count it's frequency.

Fifth Task is to Write a Python program to add 'ing' at the end of a given string and If the given string already ends with 'ing' then add 'ly' instead

So, I just check the last 3 characters of the string if they match the condition, i add ing or ly according to the condition.

Sixth Task is to Write a Python program to add 5 seconds with the current time.

So, I use the datetime python library to compute the time now and add 5 seconds to compute the time after 5 seconds using timedelta method.

Seventh Task is to compute the value of P at point x, where P is coefficients of a polynomial.

So, I have used numpy array to take float coefficients from user and split them in addition to the value of x, and then use the numpy polyval method to compute the value of P

Eighth Task is to compute the determinant of a given NxN matrix

So, I have used input function to take the matrix rank from user in addition to the values of the matrix and split them according to the matrix rank, and use the numpy det method to compute the determinant.

Ninth Task is to design and build a simple GUI Calculator.

So, I have used python tkinter library to design and build it.

- Create click button event to make action when clicking any button.
- Create clear function to clear output on calc screen
- Initialize calc frame using grid
- Create all buttons with their labels using TK button and label methods

Second one: Digit-Recognizer Model

In this model, we use MNIST dataset to identify the handwritten digits

1. I load the required libraries to prepare and preprocess the data
2. I Check for null and missing values
3. I have normalize the data to be between 0 and 1 to prevent overfitting the model
4. Reshape image in 3 dimensions (height = 28px, width = 28px , canal = 1)
5. Label Encoding, encode labels to one hot vectors (ex : 2 -> [0,0,1,0,0,0,0,0,0])
6. Split the train and the validation set for the fitting
7. Build the model using CNN with
 - MaxPooling of 2
 - Dropout of 0.25

- # of filters from 32 after first 2 lay to 64 filters in the third and fourth layer
 - Use Relu activation function
 - Then add Flatten and Dense layer with Softmax activation
8. Use RMSprop Optimizer, categorical cross entropy loss function, and accuracy matrix.
 9. Data augmentation to prevent overfitting
 10. After Evaluate the model, i got
 - Training Accuracy: 0.975
 - Validation Accuracy: 0.98

Third one: Fake and Real news Detection Model

In this model, we use two csv files of data, one for true news and the other for fake news

So, I have build this model to be able to classify real news from fake news

1. Import the required libraries and load the datasets
2. Creating check on both dfs that will be the target feature which are the true label (True or Fake)
3. Combine both dfs to be one dataset.
4. join title, text and subject to create the article feature.
5. Creating the final Dataframe with article and check.
6. Convert article to lowercase.
7. Removing punctuation.
8. Removing stop words using stopwords from nltk library.
9. Use 20% of the data to test the model
10. Use Support Vector Machine
11. After Evaluate the model, i got
 - Training Accuracy: 0.99

