

Kennesaw State University

CS 7357 Neural Nets and Deep Learning

Project 1

Description

In this project, we will implement KNN using Python. You are given three data sets:

| data type | have a label | function |
|----------------|--------------|--|
| training set | Yes | Used to train the model or determine the model parameters, such as the determination of the weights in KNN |
| validation set | Yes | Used to determine the network structure/parameter or control the complexity of the model and modify the model. |
| test set | No | Used to test the performance of the optimal model. |

A typical division is that the training set accounts for 50% of the total sample, and the others account for 25%, and all of them are randomly selected from the sample.

Dataset:

| Document number | The sentence words | Emotion |
|-----------------|--------------------------------------|---------|
| Train 1 | I buy an apple phone | happy |
| Train 2 | I eat the gig apple | happy |
| Train 3 | The Apple products are too expensive | sad |
| Train 4 | My friend has an apple | ? |

1, use KNN for classification problems. You can use one-hot matrix to represent the sentences. On the verification set, adjusting the K value and selecting different distances (Manhattan distance vs. Euclidean distance) to get a model with the best accuracy. Write the process in the experimental report.

Example:

```
k = 5 Validation set correct rate: 0.4212218649517685
k = 6 Validation set correct rate: 0.40836012861736337
k = 7 Validation set correct rate: 0.40836012861736337
k = 8 Validation set correct rate: 0.40836012861736337
```

P.S., Distance formula:

Manhattan distance: In a plane with p1 at (x1, y1) and p2 at (x2, y2), it is $|x_1 - x_2| + |y_1 - y_2|$.

Euclidean distance: In a plane with p1 at (x1, y1) and p2 at (x2, y2), it is $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

2, apply the model parameters (K, distance type, etc.) obtained in step 1 on the test set, and save the output result as "my_result.csv".

Example:

my_result

| Words (split by space) | label |
|--|-------|
| senator carl krueger thinks ipods can kill you | joy |
| who is prince frederic von anhalt | joy |
| prestige has magic touch | joy |
| study female seals picky about mates | joy |
| no e book for harry potter vii | joy |
| blair apologises over friendly fire inquest | fear |

Submission

You have to submit the followings to D2L:

1. MS word file
 - Describe what you have done for the homework assignment.
2. Python source code file(s)
 - Must be well organized (comments, indentation, ...)
 - You need to upload the "original python file (*.py)" and also its "PDF" version.
 - o For the PDF file, you can just convert the source file to PDF. One way is to print the source file and save to "PDF".

You have to submit the files SEPERATELY. DO NOT compress into a ZIP file.