Mir Umar Hussain

Umarmir425@gmail.com

Section 1

"""

Created on Wed Feb 9 14:19:12 2022

@author: Mir Umar

"""

```
def LongestSubstr(s):
    def chk(start, end):
        chars = [0] * 128
        for i in range(start, end + 1):
            c = s[i]
            chars[ord(c)] += 1
            if chars[ord(c)] > 1:
                return False
        return True

    n = len(s)
```

```
res = 0
for i in range(n):
    for j in range(i, n):
        if chk(i, j):
            res = max(res, j - i + 1)
return res
```

#Time Complexity: O(N^3), where N is the length of the string.

#Space Complexity: O(min(N, M)), as HashSet is used. N is the length of the string and M is the size of the substrings.

Section 3

Q2.

K-Nearest Neighbors (K-NN)

k-NN is a supervised algorithm used for classification. What this means is that we have some labeled data upfront which we provide to the model for it to understand the dynamics within that data i.e. train. It then uses those learnings to make inferences on the unseen data i.e. test. In the case of classification this labeled data is discrete in nature.

Steps

1. Decide on your similarity or distance metric.
2. Split the original labeled dataset into training and test data.
3. Pick an evaluation metric.
4. Decide upon the value of k. Here k refers to the number of closest neighbors we will consider while doing the majority voting of target labels.
5. Run k-NN a few times, changing k and checking the evaluation measure.
6. In each iteration, k neighbors vote, majority vote wins and becomes the ultimate prediction
7. Optimize k by picking the one with the best evaluation measure.

The complexity of the algorithm is as follows:
Train: O(dim)

Test: O(n*dim)

Where:

dim: is the dimensions or the features of the training examples

n: is the number of training examples

k-Means

k-Means is an unsupervised algorithm used for clustering. By unsupervised we mean that we don't have any labeled data upfront to train the model. Hence the algorithm just relies on the dynamics of the independent features to make inferences on unseen data.

Steps

1. Initially, randomly pick k centroids/cluster centers. Try to make them near the data but different from one another.
2. Then assign each data point to the closest centroid.
3. Move the centroids to the average location of the data points assigned to it.
4. Repeat the preceding two steps until the assignments don't change, or change very little.

The complexity of the algorithm is as follows:
O(n*k*i*d)

Where:

n: is the number of points

k: is the number of clusters

i: is the number of iterations

d: is the number of attributes

Q3: L1 and L2 regularization:

Regularization is used in machine learning to avoid overfitting of data by the model. In machine learning, there are two types of regularization that are mainly used, they are L1 regularization and L2 regularization.

The main intuitive difference between the L1 and L2 regularization is that L1 regularization tries to estimate the median of the data while the L2 regularization tries to estimate the mean of the data to avoid overfitting.

A regression model that uses L1 regularization technique is called Lasso Regression and model which uses L2 is called Ridge Regression.

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds “absolute value of magnitude” of coefficient as penalty term to the loss function.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Ridge regression adds “squared magnitude” of coefficient as penalty term to the loss function. Here the highlighted part represents L2 regularization element.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Section 4:

1.
2. (B) Recurrence is T(n) = T(n-1) + O(n) and time complexity is O(n^2)

3. (B) Process does not have sufficient memory to execute
4. (B) TCP/IP
5. (C) In CSMA/CA Frame Acknowledge is used to aviod collisions