CS 559: Machine Learning: Fundamentals and Applications

Due: 2/27/2024 Wednesday 11:59 p.m.

- The assignment must be individual work and must not be copied or shared. Any tendency to cheat/copy evidence will lead to a 0 mark for the assignment.
- Students must only use Pandas, NumPy, and Spacy if the coding problem does not specify libraries/packages. Use of other libraries than specified will be penalized.
- All problems must be submitted in a single notebook file.

1 Naive Bayes Classification [40 pts]

Use the following code to generate the train data set. The code will generate a random data set with four features and classes.

```
from sklearn import datasets
X, y = datasets.make_blobs(n_samples = 400, n_features= 5,
    centers = 4, cluster_std= 2, random_state= 100)
```

a. [5 pts] Compute the prior probability of each class, $p(C_k)$.

b. [10 pts] Compute the likelihood $p(\mathbf{X}|C_k)$.

c. [15 pts] Compute the posterior probability of each point $p(C_k|\mathbf{X})$. Assign the class ID to each point.

d. [5 pts] Construct the confusion matrix to show the classification rate.

e. [5 pts] Classify the target using **sklearn.native_bayes.GaussianNB**. Report the accuracy of the model.