Assignment 3

Start Assignment

- Due Friday by 11:59pm
- Points 100
- Submitting a file upload
- File Types pdf

In this homework, we will use the Cancer dataset.

Note: You can use the built-in function from ML libraries for gradient descent, training, and validation.

Problem 1 (25pts):

(i) (7.5+2.5+2.5 points) Use the cancer dataset to build a logistic regression model to classify the type of cancer (Malignant vs. benign). First, create a logistic regression that takes all 30 input features for classification. Use 80% and 20% split between training and evaluation (test). Make sure to perform proper scaling and standardization before your training. Also, report your results, including accuracy, precision, and recall. Plot the confusion matrix representing your binary classifier. **See this notebook (https://github.com/Farah-Deeba-UNCC/Introduction-to-ML/blob/3d8984507b0babaa655bfeaa331f7a4aa92ab4e4/Notebooks/10-LogisticRegression.ipynb) for guidance regarding implementation of logistic regression.**

(ii) (7.5+2.5+2.5 points) How about adding a weight penalty here, considering the number of parameters? Add the weight penalty and repeat the training and report the results.

Problem 2 (25pts):

(i) (13 points) Use the cancer dataset to build a naive Bayesian model to classify the type of cancer (Malignant vs. benign). Use 80% and 20% split between training and evaluation (test).

(ii) (7+5) Plot your classification accuracy, precision, recall, and F1 score. Explain and elaborate on your results, comparing your results against the logistic regression classifier you did in Problem 1. **See this notebook (https://github.com/Farah-Deeba-UNCC/Introduction-to-ML/blob/3d8984507b0babaa655bfeaa331f7a4aa92ab4e4/Notebooks/11-NaiveGaussianBayes.ipynb) for guidance regarding implementation of logistic regression.**

Problem 3 (25pts):

(i) (13 points) Use the cancer dataset to build an SVM classifier to classify the type of cancer (Malignant vs. benign). Use 80% and 20% split between training and evaluation (test).

(ii) (7+5) Plot your classification accuracy, precision, recall, and F1 score. Explain and elaborate on your results, comparing your results against the classifiers you did in Problem 1 and 2. You can modify your

Naive Bayes Classifier by writing the following:

from sklearn.svm import SVC

classifier = SVC(kernel='linear', C=1.0)

Problem 4 (25pts):

(15 points) Use the cancer dataset to build a logistic regression model to classify the type of cancer (Malignant vs. benign). Use the PCA feature extraction for your training. Perform N number of independent training (N=1, …, K). Identify the optimum number of K, principal components that achieve the highest classification accuracy.

(5 + 5 points) Plot your classification accuracy, precision, recall, and F1 score over a different number of Ks. Explain and elaborate on your results and compare them against problems 1 and 2.

Accessing the cancer dataset:

The cancer dataset may be obtained through the sklearn library as shown below:

4 Breast Cancer dataset read

[38]:
```
import pandas as pd
```

[59]:
```
from sklearn.datasets import load_breast_cancer
breast = load_breast_cancer()
X = breast.data
print(X.shape)
Y = breast.target
```

```
(569, 30)
```

[60]:
```
breast_input = pd.DataFrame(X)
breast_input.head()
```

[60]:

	0	1	2	3	4	5	6	7	8	\
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419	
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812	
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069	
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597	