**PART A**

**EXPERIMENT NO. 8**

**A.1 AIM: -** TO detect Breast Cancer Diagnosis via Logistic Regression

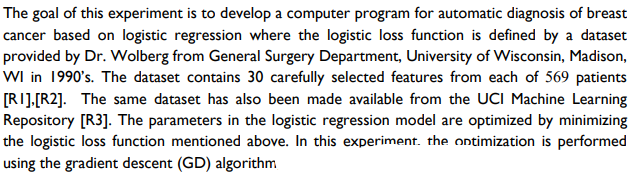
**A.2 Prerequisite**

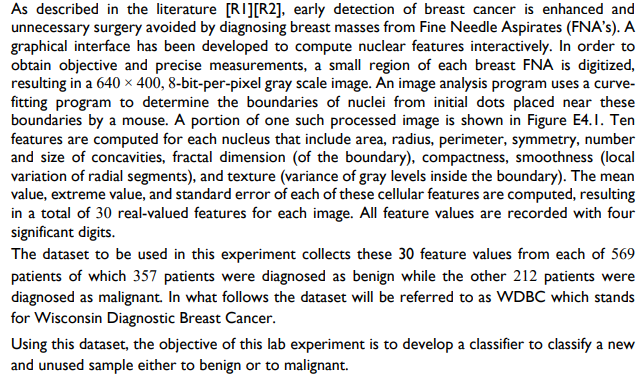
* Different programming language (Python or Java), Understanding of Machine Learning Algorithms, Machine Learning Algorithms

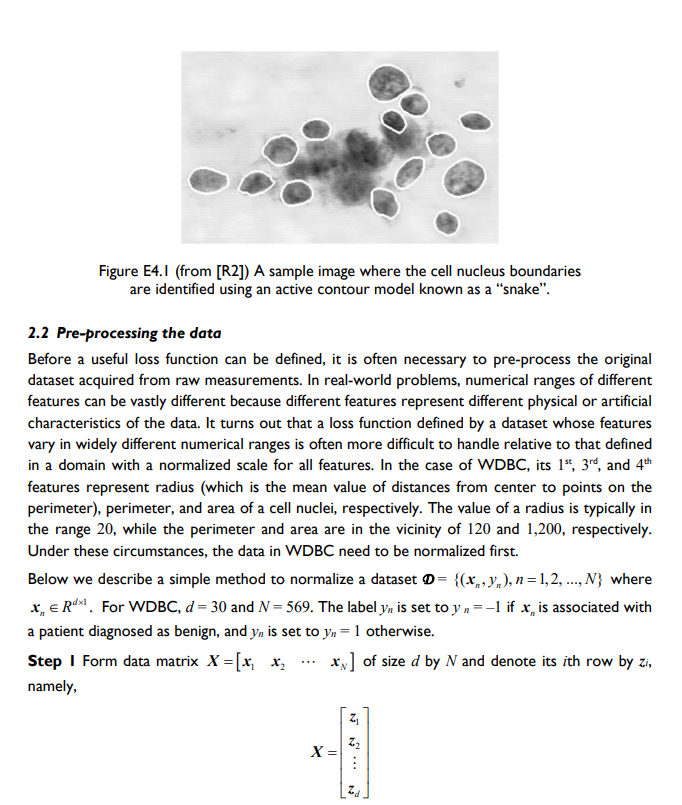
**A.3 Outcome**

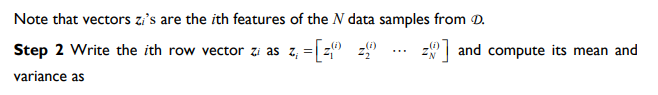
After successful completion of this experiment students will be able to Optimize the problem.

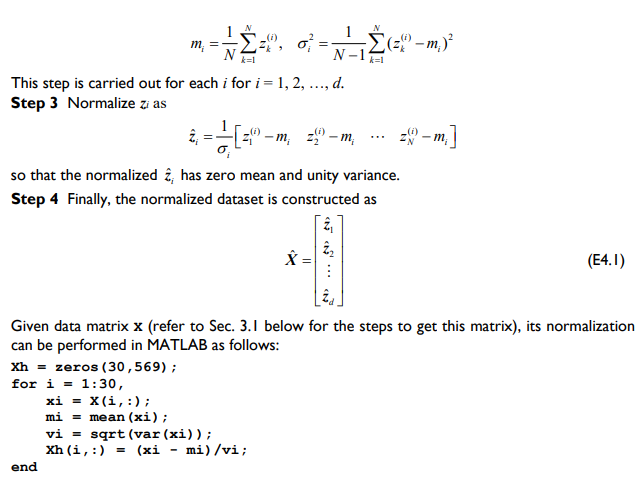
**A.4 Theory**



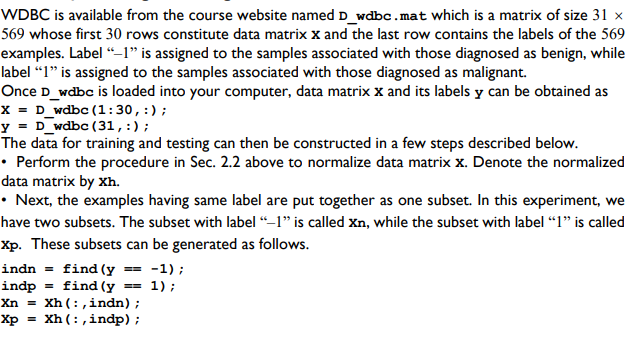


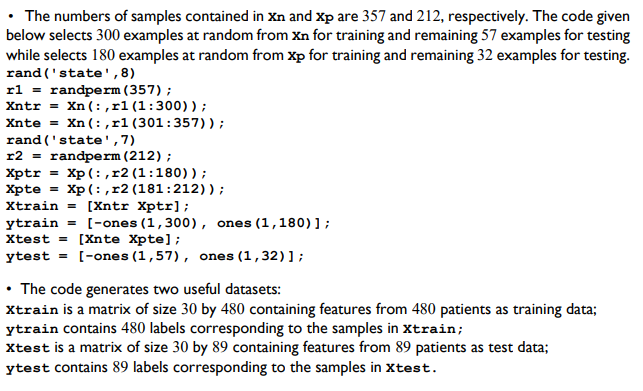


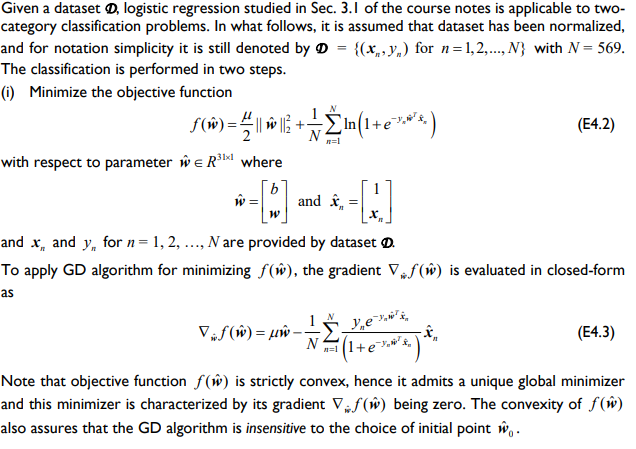


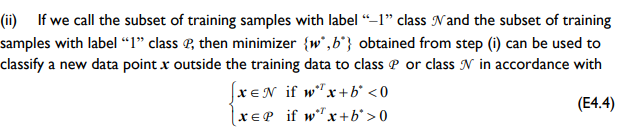


Procedure:









Links: <https://archive.ics.uci.edu/ml/datasets/breast%2Bcancer%2Bwisconsin%2B(Prognostic)>

Or

<https://www.kaggle.com/datasets/yasserhessein/dataset-breast-cancer-wisconsin>

PART B

|  |  |
| --- | --- |
| Roll No. C015 | Name: Prachi Dave |
| Class : BTI SEM 10 | Batch : EB1 |
| Date of Experiment: 22/03/24 | Date of Submission |
| Grade : |  |

**B.1 Documentation written by student:**

import pandas as pd

import numpy as np

from sklearn.model\_selection import train\_test\_split, GridSearchCV

from sklearn.linear\_model import LogisticRegression

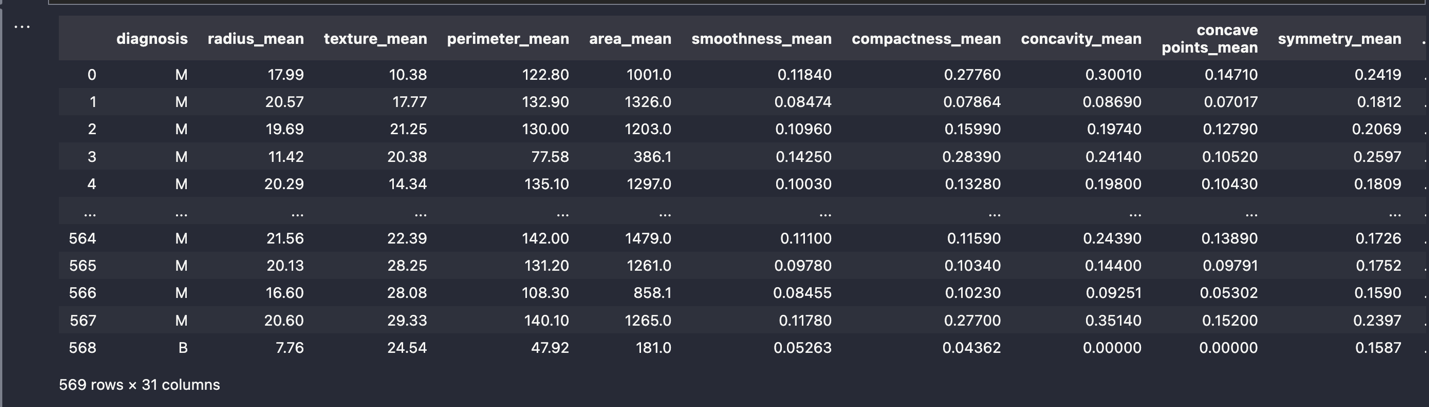
from sklearn.preprocessing import StandardScaler

from sklearn.metrics import classification\_report

df = pd.read\_csv('data.csv')

df.drop(*columns*=["id", "Unnamed: 32"], *inplace*=True)

df



X = df.iloc[:, 1:]

y = df.iloc[:, 0]

y = np.where(y == 'B', 0, 1)

scaler = StandardScaler()

X = scaler.fit\_transform(X)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, *random\_state*=42, *test\_size*=0.2)

param\_grid = {

"penalty": ["l1", "l2"],

"C": [0.1, 1, 10],

"solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"]

}

grid\_search = GridSearchCV(LogisticRegression(), *param\_grid*=param\_grid, *cv*=5)

grid\_search.fit(X\_train, y\_train)

best\_log\_reg = grid\_search.best\_estimator\_

preds\_optimized = best\_log\_reg.predict(X\_test)

accuracy\_optimized = best\_log\_reg.score(X\_test, y\_test)

print("Accuracy of Optimized Logistic Regression:", accuracy\_optimized)

print("Classification Report for Optimized Logistic Regression:\n", classification\_report(y\_test, preds\_optimized))

A screenshot of a logistic report

Description automatically generated

**B.2 Observations and learning:**

The optimized logistic regression model achieves an impressive accuracy of approximately 99.12%. Upon examining the classification report, it's evident that the model performs exceptionally well across both classes (0 and 1), with high precision, recall, and F1-score values. Specifically, for class 0, it achieves near-perfect precision, recall, and F1-score of 0.99, indicating precise and accurate predictions for benign cases. For class 1, although the precision is perfect (1.00), there is a slightly lower recall (0.98), suggesting a small number of malignant cases might have been misclassified as benign.

**B.3 Conclusion:**

In summary, the optimized logistic regression model demonstrates remarkable performance in accurately classifying breast cancer cases. Its high precision, recall, and F1-scores, coupled with an overall accuracy exceeding 99%, underscore its reliability and effectiveness in identifying both benign and malignant tumors. This underscores the importance of fine-tuning model parameters to achieve optimal performance, especially in critical applications like medical diagnosis.