CSE4088 Introduction to Machine Learning

Diagnosis Prediction and Classification of Breast Cancer

Team Members:

Yunus Ahmed Stahlschmidt 150119814

Sameeh N O Kunbargi 150119693

Muhammed Fatih Öztel 150119907

Abstract

Breast cancer is one of the most common cancers among women worldwide, representing the majority of new cancer cases and cancer-related deaths according to global statistics, making it a significant public health problem in today's society.

This analysis aims to observe which features are most helpful in predicting malignant and benign breast cancer. Furthermore, our in-depth analysis will allow us to observe general trends which may aid us in the model and hyperparameter selection. The goal is to classify whether the cancer is benign or malignant. To achieve this we will use different machine learning classification methods to fit a function to our data that can predict the class of new input.

What we have accomplished so far:

As we mentioned in the project plan, We did an initial literature search. We found some articles about the diagnosis prediction and classification of breast cancer. We searched and found what kind of Machine Learning algorithms are used for this problem. We decided to implement the following machine learning algorithms:

- Perceptron learning
- Logistic regression
- Linear regression classifier
- Neural Networks

Remaining work:

Until now, we did a literature search, decided on the algorithms, and inspected the dataset, but further in-depth inspection of the dataset is needed. Now, we need to start the implementation of these algorithms. First of all, must parse our data in a format that will allow us to make use of vectorization techniques in order to drastically improve the performance of the algorithms, as long as we don't apply methods like Stochastic Gradient Descend. For the first 3 sub-tasks, we have sufficient experience thanks to the course homework, but in order to implement a neural network for this problem, we have to conclude further research.

Revised Project plan:

We revised our project plan and since our literature search took much more time than we expect, we couldn't start a basic implementation. Our new revised project plan is as follows:

- Searching the project subject: By the 5th week, We decided on the project topic that we will work on. We prepare related project plan documents until November 9th
- Initial literature search: After the project topic is approved, we will start with the literature search. Check related breast cancer articles and find what kind of Machine Learning algorithms are used.
- Preparing the midterm report: by November 30th, Review the dataset. Prepare the midterm report and decide what algorithms to implement.
- Preparing final report & final presentation: by January 18th, Implement all algorithms, compare all algorithms implemented, Report the final results and prepare a presentation.

List of references:

Harinishree, M. S., Aditya, C. R., & Sachin, D. N. (2021, April). Detection of Breast Cancer using Machine Learning Algorithms–A Survey. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1598-1601). IEEE.

SAYGILI, A. (2018). Classification and diagnostic prediction of breast cancers via different classifiers. *International Scientific and Vocational Studies Journal*, *2*(2), 48-56.

Telsang, V. A., & Hegde, K. (2020, December). Breast Cancer Prediction Analysis using Machine Learning Algorithms. In *2020 International Conference on Communication, Computing and Industry 4.0 (C2I4)* (pp. 1-5). IEEE.

https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3 Accessed 06 Nov 2021

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html Accessed 06 Nov 2021