Binary Classifier Using SVM to Classify Breast Cancer

Brendan Veit

**Abstract**—A SVM Machine Learning Algorithm to determine whether a Breast Cancer tissue sample is Benign or Malignant.

**Index Terms**—Breast Cancer, Machine Learning, Support Vector Machine, Data Analysis

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

————————————————

—————————— ◆ ——————————

# 1 Introduction

T

his project was to create a Machine Learning algorithm to predict if Breast Cancer tissue is benign or malignant. The algorithm that was used to classify the data for this program was a Support Vector Machine Algorithm. An analysis of which features contained more weight in either a benign or malignant classification.

# 2 Data

T

he data that was used to teach the SVM algorithm is the [Breast Cancer Wisconsin (Diognostic) Data Set](https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data). The dataset contains thousands of breast cancer tissue samples with 30 different features as well as the classifer (benign [B] or malignant [M]) for that tissue sample. The data is incredibly clean with no missing data, but the “id” column had to be removed when the data was processed by the program.

# 3 Outcomes

T

he program generally runs between 96% and 100% accuracy with predicting weather a tissue sample is benign or malignant. The program also generated a graph (see below) based on the Coefficient Magnitude of the features in the SVM algorithm. In the context of a linear SVM classifier, the coefficients represent the weights assigned to each feature. The sign (positive or negative) of the coefficient indicates the direction of the relationship between the feature and the predicted class.

1. Positive Coefficient: A positive coefficient means that as the value of the feature increases, the odds of the predicted class being positive (in this case, Malignant) also increase.
2. Negative Coefficient: A negative coefficient means that as the value of the feature increases, the odds of the predicted class being negative (in this case, Benign) increase.

# 4 Conclusions

The SVM Classifier was able to determine with over 96% accuracy whether a tissue sample was Benign or Malignant based on 30 different features presented to it. The SVM Classifier also determined which features led to a Benign or Malignant classification. Both radius\_mean and texture\_se led to more Benign assignments, while compactness\_worst, concavity\_worst, concave points\_worst, and symmetry\_worst all led to more Malignant classifications.

**Acknowledgment**

Thank you to Professor Chakraborty for teaching me the fundamentals of Machine Learning that were used to complete this project.

A graph with different colored bars

Description automatically generated