Pune Institute of Computer Technology
Dhankawadi, Pune

Breast Cancer Classification Using Support Vector Machine

SUBMITTED BY

Rajwinder Singh (41152)
Sahil Singh (41155)
Samiksha Jagadale (41156)
Sanchit Raina (41157)

Under the guidance of
Prof. Rutuja A. Kulkarni

DEPARTMENT OF COMPUTER ENGINEERING
Academic Year 2021-22

INTRODUCTION

Breast cancer is the most common cancer amongst women in the world. It accounts for 25% of all cancer cases, and affected over 2.1 Million people in 2015 alone. It starts when cells in the breast begin to grow out of control. These cells usually form tumors that can be seen via X-ray or felt as lumps in the breast area.

Early diagnosis significantly increases the chances of survival. The key challenges against it's detection is how to classify tumors into malignant (cancerous) or benign(non cancerous). A tumor is considered malignant if the cells can grow into surrounding tissues or spread to distant areas of the body. A benign tumor does not invade nearby tissue nor spread to other parts of the body the way cancerous tumors can. But benign tumors can be serious if they press on vital structures such as blood vessels or nerves.

Machine Learning technique can dramatically improve the level of diagnosis in breast cancer. Research shows that experienced physicians can detect cancer by 79% accuracy, while a 91 %(sometimes up to 97%) accuracy can be achieved using Machine Learning techniques.

OBJECTIVE

To classify tumors into malignant (cancerous) or benign (non-cancerous) using features obtained from several cell images.

H/W AND S/W REQUIREMENTS

Hardware Requirements: PIV, 2GB RAM, 500 GB HDD

Software Requirements : Anaconda with Python 3.7, Jupyter Notebook

THEORY CONCEPTS

A Support Vector Machine (SVM) is a binary linear classification whose decision boundary is explicitly constructed to minimize generalization error. It is a very powerful and versatile Machine Learning model, capable of performing linear or nonlinear classification, regression and even outlier detection.

SVM is well suited for classification of complex but small or medium sized datasets.

The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

The disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation

RESULTS

```
print(classification_report(y_test,y_predict))
```

```
              precision    recall  f1-score   support

         0.0       0.91      1.00      0.95        48
         1.0       1.00      0.92      0.96        66

    accuracy                           0.96       114
   macro avg       0.95      0.96      0.96       114
weighted avg       0.96      0.96      0.96       114
```

CONCLUSION

By performing this miniproject we were able to understand the basics of SVM classifier. We also trained an SVM model to make accurate breast classifications.