

BREAST CANCER PREDICTION

(sheema7)

1. Introduction

Breast cancer is a prevalent and potentially life-threatening disease affecting millions of individuals worldwide. Early detection and accurate classification of tumors play a crucial role in improving patient outcomes. In this documentation, we explore the application of Support Vector Machines (SVM) for the classification of breast cancer based on features extracted from diagnostic imaging.

2. Data Loading and exploration

Importing necessary libraries:

```
import numpy as np

import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler

from sklearn.svm import SVC

from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

Loading and Previewing the dataset:

```
data = pd.read_csv('data.csv')

print(data)
```

Dataset Information and Descriptive Statistics:

```
print(data.info())

print(data.head())

print(data.tail())

print(data.describe())

print(data.shape)
```

Handling Missing Values:

```
missing_val = data.isnull().sum()

print(missing_val)
```

Target Variable Distribution:

```
data['diagnosis'].value_counts()
```

3. Data Splitting and Standardization

Splitting the dataset into training set and testing set:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state = 40, test_size = 0.5)
```

Standardizing Feature Data:

```
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

4. Support Vector Machines(SVM)

Support Vector Machines are powerful supervised learning models used for classification and regression tasks. In this project, we focus on the classification aspect.

Model Initialization and Training:

```
from sklearn.svm import SVC
model = SVC(kernel='linear')
model.fit(X_train, y_train)
```

Confusion Matrix:

```
y_pred = model.predict(X_test)

print(confusion_matrix(y_test, y_pred))
```

Classification Report:

```
Print(classification_report(y_test, y_pred))
```

Accuracy Calculation:

```
acc = accuracy_score(y_test, y_pred)

print(f"Accuracy of SVM model: {acc:.2f}")
```

5. Conclusion

In this documentation, we explored the application of Support Vector Machines for the classification of breast cancer based on diagnostic imaging features. The dataset was loaded, explored, and pre-processed. The SVM model was trained, evaluated, and its performance metrics were discussed.

Summary of Findings:

The SVM model demonstrated promising accuracy in classifying breast cancer. The confusion matrix and classification report provide insights into the model's performance, showing its ability to differentiate between benign and malignant tumors.

Limitations and Future Work:

While the SVM model shows promise, there are limitations to consider. Future work could involve exploring different kernel functions, optimizing hyperparameters, and incorporating additional feature engineering techniques to enhance model performance.

Acknowledgments:

We acknowledge the dataset source and the contributions of the research community in breast cancer classification. Their work laid the foundation for our exploration and understanding of this critical area in healthcare.

References:

Include citations to relevant papers, articles, or datasets used in the project.