

Breast Cancer Classification Using Neural Networks

Project Overview

This project implements a neural network based classification system to predict whether a breast tumor is malignant or benign. The solution is built using Python machine learning libraries and a feedforward neural network implemented with Keras. The objective of the project is to demonstrate an end to end supervised learning pipeline including data loading preprocessing model training evaluation and prediction.

Dataset Description

The dataset used in this project is the Breast Cancer Wisconsin dataset available through sklearn datasets. It contains diagnostic measurements computed from digitized images of breast mass tissue samples. Each sample consists of thirty numerical features representing properties such as radius texture perimeter area smoothness compactness concavity symmetry and fractal dimension. The target variable indicates tumor type where zero represents malignant and one represents benign.

Data Exploration and Preprocessing

The dataset is converted into a Pandas DataFrame to enable easier exploration and analysis. Basic exploratory checks such as viewing dataset shape column information and missing values are performed. The target labels are appended as a separate column. The dataset contains no missing values and all features are numerical which simplifies preprocessing. The data is then split into training and testing sets to evaluate model generalization.

Feature Scaling

StandardScaler is applied to normalize the input features. Feature scaling is crucial for neural networks as it ensures that all input features contribute equally to learning. The scaler is fitted only on the training data to prevent data leakage and then applied to both training and testing sets.

Neural Network Architecture

The neural network is built using the Keras Sequential API. The architecture consists of an input flattening layer followed by a hidden dense layer with ReLU activation. The output layer uses softmax activation to produce class probabilities for benign and malignant tumor classes. This architecture enables the model to learn non linear patterns present in medical diagnostic features.

Model Compilation

The model is compiled using the Adam optimizer which adapts learning rates during training for faster convergence. Sparse categorical crossentropy is used as the loss function since the target labels are integer encoded. Accuracy is selected as the evaluation metric to measure classification performance during training and validation.

Model Training

The neural network is trained on the scaled training data using multiple epochs. A portion of the training data is reserved for validation to monitor overfitting. During each epoch the model updates its weights through backpropagation and gradient descent. Training and validation accuracy trends provide insight into learning behavior.

Model Evaluation

After training the model is evaluated on the test dataset. Predicted probabilities are converted into class labels using argmax. The final accuracy score indicates the effectiveness of the neural network in classifying tumors. The model demonstrates strong predictive performance suitable for medical classification tasks.

Predictive System

A predictive system is implemented by passing new patient feature values through the trained scaler and neural network. The system outputs whether the tumor is malignant or benign based on learned patterns. This demonstrates how trained models can be used for real world inference.

Key Takeaways

This project demonstrates a complete neural network classification workflow applied to a real medical dataset. It highlights the importance of preprocessing feature scaling model design and proper evaluation. The solution shows how neural networks can effectively support medical diagnosis and decision making.

Author

Satyam Gajjar