



Advanced Machine Learning Documentation

MODELS

SVR : REGRESSION

DECISION TREE : CLASSIFICATION

NEURAL NETWORK : CLASSIFICATION

SVR

DATASET

| DATASET NAME | *HOUSE PRICES ADVANCED REGRESSION TECHNIQUES |
|-----------------------|---|
| TRAIN | %80 |
| TEST | %20 |
| VALIDATION | %20 |
| FINAL FEATURES | 46 |
| FOLDS | 8 |

IMPLEMENTATION

Data Collection and Exploration

The code begins with data collection from a Kaggle dataset and explores the dataset using descriptive statistics, visualizations, and correlation analysis

Data Preprocessing and Feature Engineering:

The preprocessing steps include handling missing values, encoding categorical variables (using both ordinal and one-hot encoding), and feature engineering to create new relevant features. The feature engineering part intelligently creates new features like 'HOUSEAGE', 'HOUSEREMODELAGE', 'TOTALSF', 'TOTALAREA', 'TOTALPATHES', and 'TOTALPORCHSF' which can potentially improve model performance.

Model Training and Evaluation

The code trains two regression models: linear regression and support vector regression (SVM). Both models are evaluated using metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared on a test set. This provides a comprehensive assessment of model performance. - Additionally, cross-validation is performed to evaluate the SVM model, providing insights into its generalization performance.

Hyperparameter tuning

done through techniques like grid search , can significantly improve the performance of machine learning models by finding the optimal set of hyperparameters. it can help fine-tune the SVM model and potentially improve its predictive accuracy.

Model Comparison and Visualization

The code concludes with a visualization comparing the performance of linear regression and SVM models using bar plots, providing a clear comparison of their scores.

Decision Tree

DATASET

| DATASET NAME | MALARIA SET IMAGES DATASET |
|----------------------------|---|
| NUMBER OF CLASSES | 2 |
| LABELS | PARASITIZED , UNINFECTED |
| TOTAL NUMBER OF SAMBLES | 27,058 |
| THE SIZE OF IT | IT DYFFERS BASED ON THE SIZE OF IMASGE . KB3:18 IT IS IN RANGE |
| NUM SAMPLE USE IN TRAINING | %70 |
| NUM SAMPLE USE IN TESTING | %30 |

FEATURE EXTRACTION

the number of features extracted per image is 2916

IMPLEMENTATION

HYPERPARAMETERS

The hyperparameters used in this Decision Tree classifier are criterion is "GINI" MAX_DEPTH is (15) (10)MIN_SAMPLES_SPLIT MIN_SAMPLES_LEAF (2)

CROSS_VALIDATION

5 folds and the ratio of training/validation is splits the data into 80% training and 20% validation sets

NEURAL NETWORK

DATASET

| DATASET NAME | MALARIA IMAGES DATASET |
|----------------------------|-----------------------------------|
| NUMBER OF CLASSESS | 2 |
| THEIR LABELS | PARASITIZED , UNINFECTED |
| TOTAL NUMBER OF SAMPLES | 27,472 |
| THE SIZE OF IT | 3:18IT IS IN RANGE |
| NUM SAMPLE USE IN TRAINING | 80%WITH FEATURE EXTRACTION |
| NUM SAMPLE USE IN TESTING | %20 |
| VALIDATION | %20 |
| NUM SAMPLE USE IN TRAINING | WITHOUT FEATURE EXTRACTION 70% |
| NUM SAMPLE USE IN TESTING | %30 |
| VALIDATION | %30 |

IMPLEMENTATION

ANN MODEL WITHOUT FEATURE EXTRACTION

Number of Hidden Layers: The model has three hidden layers

Number of Neurons per Hidden Layer: The number of neurons in the hidden layers are 256, 128, and 64, respectively

Activation Functions: The activation function used in each hidden layer is the Rectified Linear Unit (ReLU), and the output layer uses the linear activation function

Loss Function: The model uses Sparse Categorical CROSSENTROPY loss function for multi-class classification problems

Learning Rate: The learning rate of the Adam optimizer is set to 0.001

Batch Size: The batch size used during training is 35

Epochs: The maximum number of epochs is set to 20

Early Stopping: Early stopping is implemented with a patience of 15 epochs and restoration of the best weights.

ANN MODEL WITH FEATURE EXTRACTION

Number of Hidden Layers: The model has three hidden layers

Number of Neurons per Hidden Layer: The number of neurons in the hidden layers are 64, 32, and the number of classes (variable), respectively

Activation Functions: The activation function used in the first two hidden layers is the Rectified Linear Unit (ReLU), and the output layer uses the SOFTMAX activation function

Loss Function: The model uses Sparse Categorical CROSSENTROPY loss function for multi-class classification problems

Optimizer: The Adam optimizer is used with a learning rate of 0.001

Batch Size: During training, a batch size of 64 is used

Number of Epochs: The model is trained for a maximum of 50 epochs

Early Stopping: Early stopping is implemented with a patience of 3 epochs, and restoration of the best weights is enabled

Early Stopping: Early stopping is implemented with a patience of 15 epochs and restoration of the best weights