Audio Classification using Multi-layer Perceptron (MLP)

Abstract

This report presents the implementation and analysis of a multi-layer perceptron (MLP) neural network for the classification of cat and dog sounds based on their Mel-frequency cepstral coefficients (MFCCs). The dataset used for training and evaluation is sourced from Kaggle, containing audio recordings of both cats and dogs. The neural network is implemented using the Keras framework

1. Introduction

Audio classification involves the categorization of audio recordings into predefined classes. This assignment specifically focuses on the classification of cat and dog sounds. The goal is to design an effective neural network model capable of accurately distinguishing between these two classes.

2. Data Collection and Preprocessing

The dataset consists of audio recordings of cats and dogs obtained from Kaggle. A balanced dataset is crucial for effective training, and class labels are extracted from the filenames. Mel-frequency cepstral coefficients (MFCCs) are computed as features for each audio file, providing a representative representation of the audio content.

3. Model Design and Architecture

The neural network architecture comprises three hidden layers with 128, 64, and 32 units, respectively, utilizing the rectified linear unit (ReLU) activation function. The output layer consists of a single unit with a sigmoid activation function for binary classification. The model is compiled using the Adam optimizer and binary crossentropy loss, suitable for binary classification tasks.

4. Dataset Splitting

The dataset is split into training and test sets with an 80-20 ratio, ensuring an adequate amount for both training and evaluation. The training set is further subdivided to include a validation set, aiding in hyperparameter tuning during training.

5. Model Training

The model is trained for 10 epochs with a batch size of 32, and the training process is monitored using accuracy metrics. The validation split allows for tracking the model's performance on an unseen subset of the training data.

6. Model Evaluation

The trained model is evaluated on the test set, yielding metrics such as loss and accuracy. Additionally, the classification report provides a detailed breakdown of precision, recall, and F1-score for each class, offering insights into the model's performance.

7. Results and Discussion

The training history plot demonstrates the model's learning progress over epochs, with both training and validation accuracy displayed. The evaluation results showcase the model's accuracy on the test set and the detailed classification report provides a deeper understanding of its performance.

8. Conclusion

In conclusion, the implemented MLP neural network demonstrates promising results in classifying cat and dog sounds based on MFCC features. The balanced dataset, effective feature extraction, and careful model design contribute to the model's success. Further experimentation, including hyperparameter tuning and additional layers, could potentially enhance performance.