Project Overview

- **Goal**: Develop a neural network model to classify music genres based on extracted audio features.
- **Dataset**: GTZAN 30-second music samples dataset.
  - Source: kaggle
  - **Samples:** 1000 tracks (10 genres, 100 tracks each).
  - **Features:** 27 audio features per sample, including:
    - Chroma features
    - Spectral features
    - Timbre
    - Tempo
- **Model Architecture:**
  - **Input layer:** 27 neurons (one per feature)
  - **First hidden layer:** 256 neurons, ReLU activation
  - **Second hidden layer:** 128 neurons, ReLU activation
  - **Output layer:** 10 neurons (for 10 genres)

Problems Faced During Training

- **Loss Plateau:** During training, the model's loss stagnated around 2.3.
  - **Cause:** Poor optimization settings and unnormalized input features.
- **Solutions Implemented:**
  1. **Feature Normalization:**
     1. Applied StandardScaler to standardize features.
     2. Result: Faster convergence and stabilized training.
  2. **Optimizer Settings:**
     1. Learning rate (lr) set to 1e-3.
     2. Momentum added (momentum=0.9) for smoother gradient updates.

Addressing Overfitting

- **Initial Problem:** Training accuracy was high (~99%), but test accuracy remained low (~65%).
- **Techniques to Improve Generalization:**
  1. **Dropout Layers:**
     1. Applied Dropout(p=0.3) after each hidden layer.
     2. Result: Reduced overfitting and increased test accuracy to **74%**.
  2. **Weight Decay:**
     1. Added L2 regularization (weight_decay=2e-4) to penalize large weights.
     2. Prevented the model from memorizing the training set.

Final Results

- **Test Accuracy:** ~74%
- **Training Time:** ~1000 epochs.
- Model was able to generalize after regularization.
- **Key Takeaways:**
  - Normalizing features significantly improved convergence.
  - Dropout combined with weight decay effectively reduced overfitting.