

Projekat: Klasifikacija instrumenata – prvi koraci

Kako smo krenuli

Dogovorili smo se da krenemo od manjeg podskupa podataka da bismo prvo proverili da sve radi. Umesto kompletnog train skupa (289k fajlova), uzeli smo test split (~4k snimaka). Cilj je bio da postavimo ceo pipeline – od podataka do prvog modela – pa da kasnije proširimo.

Šta smo uradili

1. Učitavanje podataka

- Skinuli smo i raspakovali test split NSynth dataseta.
- Učitani su metapodaci iz examples.json i napravljen DataFrame sa putanjama do WAV fajlova i labelama.
- Izbalansirali smo mini-skup tako da svaka porodica instrumenata ima približno isti broj primera.

2. Pretprocesiranje

- WAV fajlovi su pretvoreni u log-mel spektrograme pomoću librosa.
- Normalizovali smo podatke da CNN-u bude lakše da uči.
- Dimenzije: 64 mel filtera × vremenski frejmovi.

3. Dataset i DataLoader

- Napravili smo PyTorch Dataset koji vraća (spektrogram, labelu).
- Podelili smo podatke na train i validation (80/20).

4. Model

- Implementirali smo mali CNN sa nekoliko konvolucionih slojeva i jednim potpuno povezanim slojem.
- Izlaz je softmax 10 porodica instrumenata.

5. Trening

- Koristili smo Adam optimizer i CrossEntropyLoss.
- Trenirali smo 5 epoha na mini-skupu.
- Beležili smo loss i accuracy po epohi.

6. Rezultati

- Validaciona tačnost: oko 67%.
- Neke klase su dosta dobre (npr. brass recall = 1.0), dok keyboard i reed prave više problema.
- Konfuzionna matrica lepo pokazuje gde model greši.

Naredni koraci

Dalje planiramo da:

- probamo sa većim skupom (valid/train),
- treniramo duže i na jačoj arhitekturi