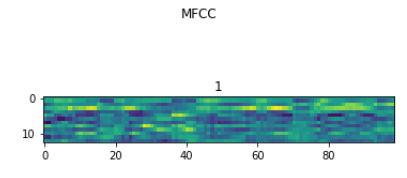
Documentatie

# Citirea si prelucrarea datelor

Prima oara am citit un sample cu scipy.io.wavfile.read pentru a afla sampling rate-ul. Apoi am citit si prelucrat datele cu ajutorul librariei librosa. Pe acestea am aplicat transformari Fourier si MFCC pentru a extrage mai multe informatii.

Pentru ca acum aveam o reprezentare vizuala a datelor m-am gandit ca un CNN ar fi optim in rezolvarea problemei. Prima oara am incercat sa antrenez modelele doar cu aceasta reprezentare, dar in final am adaugat si transformarile Fourier la feature-uri, ceea ace a condus la rezultate mai bune. De exemplu, pe datele de validare:

xgboost fara fft: 63.3%

xgboost cu fft: 72.2%

De asemenea, am observant ca foarte multe fisiere audio erau foarte silentioase. M-am gandit ca modelul nu ar avea ceva de invatat din acestea asa ca am incercat sa antrenez modelele excluzand aceste date. In general acest lucru a ajutat, dar cu un procentaj foarte mic:

svc: toate datele - 63.3%, fara date silentioase - 63.4%

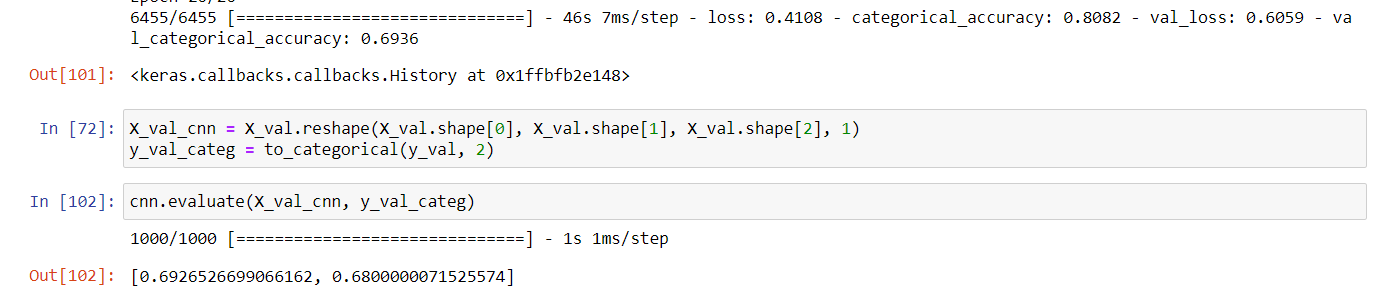
logistic regression: toate datele - 59.9%, fara date silentioase - 60.7%

mlp: toate datele - 64.1%, fara date silentioase - 64.8%

xgboost: toate datele - 63.3%, fara date silentioase - 63.1%

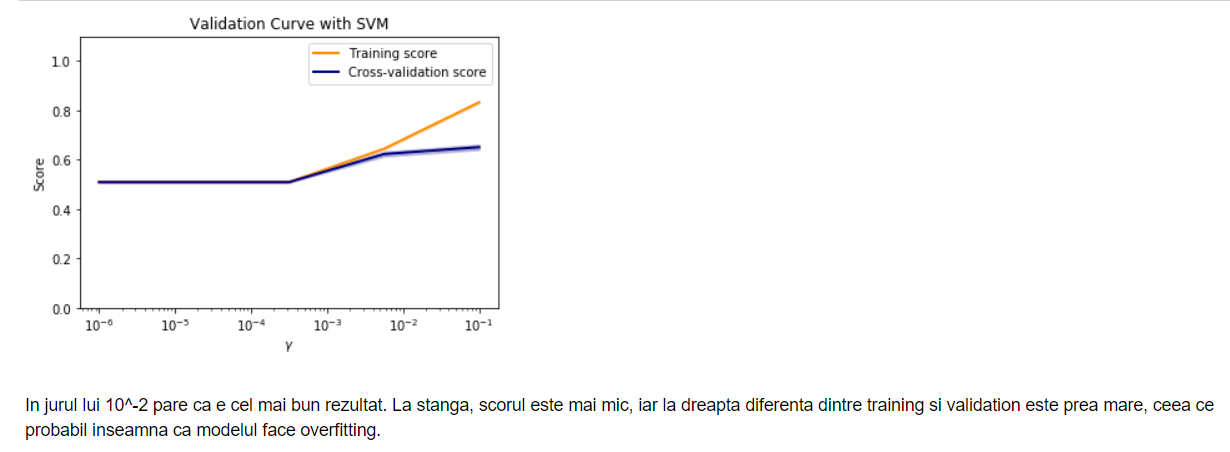
# Modele

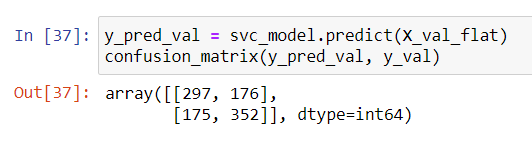
Primul model incercat a fost un CNN implementat cu keras pe MFCC. Dupa cateva rulari am observat ca modelul facea overfitting pe datele de training, asa ca o parte din ele (datele de training, pentru a putea folosi datele de validare pentru scor) i le-am lasat pentru validare in timpul antrenarii. 



Modelul a avut un scor de 68% pe datele de validare, dar nu a avut aceeasi performanta si pe cele de testare. Am incercat sa aplatizez datele si sa aplic un CNN si pe acestea, dar scorul a fost de 64.9% pe datele de validare.

Apoi am incercat un SVM pe MFCC aplatizat. Am incercat sa gasesc parametrii optimi, insa cea mai buna performanta a avut-o un SVM doar cu C-ul 10 – 64.8% pe datele din validare.





Cea mai buna performanta a avut-o modelul xgboost pe MFCC aplatizat si FFT (concatenate) cu gamma=0.1 si max\_depth=4 – 72.2% pe datele de validare. Pentru scorul final am antrenat modeul si pe datele de validare, dar cu gamma=0 si max\_depth=3 (default). Am incercat sa mai adaug si alte feature-uri, dar nu a ajutat.

La final am incercat sa folosesc CNN-ul si pe spectograma mel. Acesta a avut ca mai buna performanta dupa xgboost pe datele de test, chiar daca pe datele de validare a avut un accuracy de 63.5%.

# Clasament pe datele de validare

Xgboost – 72.7%

CNN pe MFCC– 68%

SVC – 64.9%

MLP – 64.8­%

CNN pe spectograma Mel– 63.5%

Logistic regression – 60.7%