

Razpoznavanje števk s LSTM omrežjem

Podatkovna baza:

- **Free spoken digit dataset** – baza (<https://github.com/Jakobovski/free-spoken-digit-dataset>) je angleška in vsebuje 3000 .wav posnetkov. Posneti so bili trije različni govorci, ki izgovarjajo številke (0-9). Posnetki so obrezani tako, da ni premorov. En zvočni posnetek vsebuje eno številko.
- Bazo lahko prevzamete na povezavi https://unilj-my.sharepoint.com/:u:/g/personal/mivanovska_fe1_unilj_si/ESO17dP0sWVPi2EqaAt1iVABATBuaN8uRcex_-f0zR1spg?e=nehjiN

Priprava učne in testne množice:

- Za posnetke iz podatkovne zbirke Free spoken digit dataset izračunajte MFCC značilke. V ta namen lahko uporabite Pythonovo knjižnico librosa (<https://librosa.org/doc/main/feature.html#>). Značilke uporabljajte za učenje in testiranje omrežja, ki ga boste razvili.
- 80% naključno izbranih posnetkov vzemite za učenje algoritma, preostalih 20% pa za testiranje. Pri tem bodite pozorni, da so različne številke približno enako zastopane v učni bazi.

Razvoj razpoznavalnika:

- Razpoznavalnik števk naj temelji na LSTM omrežju s polno povezano plastjo na vrhu. Število nevronov te plasti naj bo enako številu razredov baze, ki jo obravnavate (10). Ta plast ima torej vlogo razvrščevalnika značilk, ki jih LSTM omrežje izlušči. Pri razvoju modela se zgledujte po kodi, ki je na voljo na povezavi <https://github.com/moebg/spoken-digit-recognition>

Dodatni viri:

- <https://cnvrg.io/pytorch-lstm/>, <https://towardsdatascience.com/lstm-text-classification-using-pytorch-2c6c657f8fc0>

Namig: Uporabite čim več predpripravljene kode :)

