Preprocess di dati:

ho analizzato dati iniziali

ho selezionato cosi HH, EH, polarità definite, (fai qualche plot) metti statistica di quanti ne rimangono (su 10k solo 1967 tipo)

tagliato inizialmente la finestra per visualizzarne alcuni

campione di 1000 (statistica) (ho le forme d’onda ne mostro qualcuna) ho trovato tot pick con errore, tot polarità dubbie tot sbagliate totalmente

Dataset completo, su 10^6 forme d’onda, 160.000 sono velocimetri con polarità presa, utilizzo questi.

Noto alcuni sismogrammi dataset di prova hanno polarità invertite (fai alcuni esempi prendi da tracce sospette). Stimo 4% tracce abbiano polarità no buona (mislabeling, noisy label elenca tipi di rumore), temo lo stesso per il dataset maggiore. Chiedo spiegazioni a quelli dell’articolo e dicono (risposta ad Ortensia rivedi) che si, 4% è plausibile, sono messi a mano e c’è distrazione su dataset enormi (cita articolo risposta ortensia in cui anche altri dataset notevoli MNIST sono affetti da questo)

Noto che som discrimina bene tracce in cui polarità è esatta e molto evidente, errore temporale in pick è ottimo (bassissimo). Purtroppo sono esempli molto semplici, quelli meno evidenti non sono compresi qui. Provo addestrare rete su questi, che succede? Ottengo performances non molto…

Nelle classi “sospette” sono inseriti sia quelli che hanno polarità errata che quelle con ERRORE SU PICK SUPERIORE A 5 SAMPLE, ma queste ultime voglio comunque inglobarle.