

## RESULTS NNDL

- Dire che abbiamo provato restringendo le note a due ottave (adattando il codice) come suggerito dal paper, ma non avendo notato differenze abbiamo continuato con tutte le ottave
- Abbiamo fatto tuning dei parametri:
  1. learning rates iniziali
  2. parametri lambda1 e lambda 2 della regolarizzazioneconfrontando sia i risultati qualitativamente che le pch e l'errore di ricostruzione (che però era praticamente uguale per tutti), e abbiamo scelto come parametri ottimali:  
lrD = 0.00001 # Learning rate for discriminator  
lrG = 0.0005 # Learning rate for generator  
lambda1 = 0.8  
lambda2 = 10  
con 50 epoch per il training. Durante il training, la probabilità di inversione delle canzoni è tenuta fissa a p\_invert = 0.3
- Con i parametri fissi, abbiamo confrontato i risultati del modello con e senza self-attention
- Considerazioni sui risultati:
  - PCH: per la pch, durante il training non si notano miglioramenti: questo può essere dovuto al fatto che l'aspetto su cui si focalizza la pch (la frequenza di pitches) non è il focus del processo di learning, e che il modello si focalizza a migliorare altri aspetti che contribuiscono alla similarità delle canzoni. Il fatto comunque che la kl divergence non aumenta, può essere positivo perchè indica che non avviene mode collapse.  
Abbiamo notato in tutte le prove che durante il test, nella seconda metà delle canzoni la K-L divergence tra le due distribuzioni aumenta, e questo si rispecchia qualitativamente nel fatto che le canzoni degenerano.  
Nelle prime prove, quando ancora non avevamo aggiunto tutte le tecniche di regolarizzazione e i trick per depotenziare D, questo effetto era ancora più drastico perchè degeneravano in una sola nota, mentre dopo degenerano in sequenze di due o tre note che si ripetono.  
Questo succede in entrambi i casi, ma nel caso in cui la self-attention è presente quest'effetto è ridotto, e anche il valore numerico della kl divergence nel test si avvicina maggiormente a quello del training.  
grafici train-test sia con che senza self + confronto dei valori numerici
  - Embedding: Per quanto riguarda l'embedding tramite autoencoder, abbiamo confrontato solo visivamente i risultati, anche se idealmente avremmo voluto effettuare un clustering sui dati e confrontare i risultati del clustering. Alla fine abbiamo scelto di non farlo perchè non siamo riusciti a identificare un numero di cluster che fosse significativo per il tipo di dati che abbiamo  
grafici pca
  - Reconstruction loss: In entrambi i casi, così come in tutte le prove, la reconstruction loss rimane benomale la stessa e non risulta essere un indicatore significativo per confrontare

l'accuratezza dei due modelli.

valori reconstruction loss

- Valutazione qualitativa: abbiamo chiesto a n persone di valutare da 1 a 10 tre aspetti, piacevole, interessante e realistica, di un mix di canzoni tra reali, generate con attention e generate senza attention risultati penso dobbiamo fare un'immagine

FURTHER IMPROVEMENTS:

- dati migliori e più vari
- estrarre i generi per avere un numero di cluster con cui effettuare clustering, migliorando la metrica dell'autoencoder