
RICONOSCIMENTO DEI GENERI MUSICALI E DELLE LORO CARATTERISTICHE TRAMITE IL DEEP LEARNING

Baldanza Matteo^{1,2}, Masi Filippo^{1,2}, and Alessandro Zanotta^{1,2}

¹Laurea Triennale in Scienze Statistiche ed Economiche

²Studenti Magistrali in Statistica e Data Science

28 febbraio 2022

Abstract

Il progetto illustra la creazione di reti neurali di diversa tipologia che possiedono l'obiettivo di riconoscere il genere musicale di appartenenza di mille tracce musicali di trenta secondi. L'elaborato è suddiviso in due parti sulla base del tipo di metodo di classificazione. Attraverso l'utilizzo di reti convoluzionali si è cercato di catturare l'informazione presente negli spettrogrammi: immagini che mostrano, su una scala di colori, le frequenze dei suoni delle diverse tracce per risalire al relativo genere. E' stato in seguito adoperato un approccio differente: si sono estratte informazioni dalle tracce audio e raccolte in un DataSet che ha successivamente avuto la funzione di Input in reti neurali di tipo Feed Forward. In particolare per ogni traccia si sono riuscite a ottenere 164 feature relative alle caratteristiche specifiche del suono che sono diverse tra tracce di generi differenti. Per ogni rete si è effettuata un'analisi di tuning per ispezionare gli iperparametri della rete che assicurassero un ottimo adattamento ai dati e al contempo una buona performance di classificazione. Si sono inoltre testate tecniche diverse al fine di evitare il fenomeno di overfitting. I risultati delle reti hanno fornito anche un'informazione e una conferma circa le similitudini e differenze di alcuni generi musicali per quanto riguarda suoni e periodi storici. Vengono effettivamente mostrate maggiori difficoltà a classificare generi particolarmente prevalenti in diversi periodi

che hanno influenzato altre correnti musicali con suoni più standard, e invece testimoniate caratteristiche più singolari per quei generi particolari. La capacità previsiva migliore è stata esibita dalla rete neurale feed forward applicata alle tracce audio con la quale è stata ottenuta un'accuracy sui dati del test set del 90%. I risultati delle reti convoluzionali hanno invece mostrato un valore dell'accuracy nel test set pari al 58%.

1 Introduzione: Dati e Obiettivi

I dati utilizzati per le nostre analisi sono fruibili al seguente link <https://www.kaggle.com/harish24/music-genre-classification>. Sono presenti 1000 immagini che rappresentano lo spettrogramma di ogni singola traccia audio. Questo particolare diagramma, sul quale sono state raffigurate le tracce audio, è la rappresentazione grafica dell'intensità di un suono in funzione del tempo e della trasformazione logaritmica della frequenza. Sull'asse delle ascisse è riportato il tempo in scala lineare; sull'asse delle ordinate è riportata la frequenza in scala lineare o logaritmica; a ciascun punto di una data ascissa e ordinata è assegnata una tonalità di colore, rappresentante l'intensità del suono in un dato istante di tempo e a una data frequenza. Uno spettrogramma si ottiene, di solito, suddividendo l'intervallo di tempo totale (cioè quel-

*RICONOSCIMENTO DEI GENERI
MUSICALI E DELLE LORO
CARATTERISTICHE TRAMITE IL DEEP LEARNING*

lo relativo all'intera forma d'onda da analizzare) in sottointervalli uguali (detti finestre temporali) di durata da 5 a 10 ms e calcolando la trasformata di Fourier della parte di forma d'onda contenuta in ciascuna finestra, che fornisce l'intensità del suono in funzione della frequenza. Le trasformate di Fourier, relative alle diverse finestre temporali, vengono poi assemblate in base alle diverse note musicali. Si propongono qui sotto due spettrogrammi relativi a due tipi di generi musicali diversi.

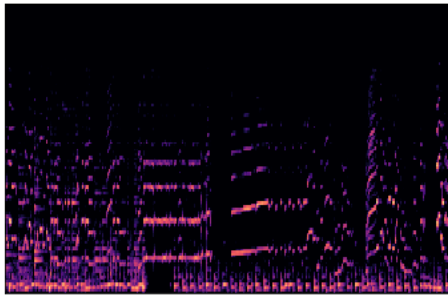


Figura 1: Spettrogramma Jazz

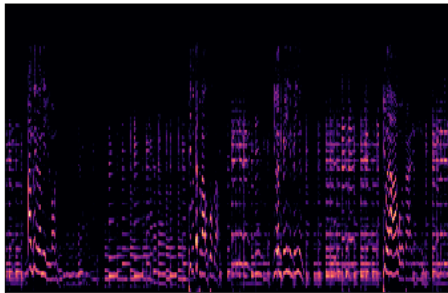


Figura 2: Spettrogramma Blues

Si può osservare come non sia banale la differenza fra le due immagini. Non appare infatti immediato capire i pattern che distinguono chiaramente i due generi. Ci si aspetta quindi una certa difficoltà da parte della rete neurale.

Si hanno poi a disposizione anche 1000 tracce musicali, di durata pari a 40 secondi. Le tracce per poter essere utilizzate all'interno di una rete neurale di tipo feed forward, sono state sottoposte a uno strumento di rilevazione delle frequenze sonore di cui si parlerà nell'apposita sezione.

L'elaborato ha come scopo principale l'individuazione della rete neurale che abbia una performance previsiva migliore. Si vuole quindi trovare il miglior metodo/modello capace di identificare un genere musicale (tra i 10 presenti) sulla base delle immagini e delle tracce. Si desidera quindi addestrare le reti neurali affinché riescano ad imparare dai dati forniti a classificare correttamente la melodia musicale ricevuta come input.

Metodi di Machine Learning e Statistical learning di questa tipologia possono tornare utile per siti, applicazioni e piattaforme musicali che hanno un bacino di brani musicali molto elevato e devono agevolare l'organizzazione di essi classificandoli nei vari generi di appartenenza.

La nostra variabile target è denominata "label" è di tipo categoriale ed è divisa in dieci etichette relative ai diversi generi musicali. Inizialmente è ragionevole monitorare la frequenza con cui ogni categoria appare all'interno delle immagini e degli audio. Il risultato è il seguente:

Blues	Classical	Country	Disco	Hip Hop
100	100	100	100	100
Jazz	Metal	Pop	Reggae	Rock
100	100	100	100	100

Come mostrato, non sono state necessarie operazioni sulla variabile risposta in quanto le categorie sono ben bilanciate nelle varie etichette e si è deciso di suddividere l'intero campione di immagini in 700,200 e 100 rispettivamente per il training set, il validation set e il test set.

2 Convolutional Neural Network

Per la classificazione delle tracce la prima procedura adoperata sono state le reti convoluzionali (CNN), come preannunciato nella sezione 1.

Le CNN sono dei particolari algoritmi che hanno un'architettura specifica e possiedono la capacità di individuare *pattern* nelle immagini con il fine di riconoscere quanto mostrato e quindi di classificare l'immagine stessa. Le reti convoluzionali si compongono di una serie di strati con compiti diversi tra di essi: i primi strati (compreso quello di input) sono i convoluzionali e permettono di utilizzare direttamente le immagini senza l'operazione di feature extraction ovvero senza che i pixel vengano srotolati all'interno dei tensori. In questa fase si va ad applicare un kernel a diverse parti di immagine formata dai pixel del colore. Attraverso l'operazione matematica di convoluzione tra kernel e sottoimmagine si ottiene un valore sintesi da ogni zona della figura che insieme agli altri risultati della operazione andrà a formare l'*activation map*. Ogni mappa permette di estrapolare informazioni diverse, in modo tale da avere una versione semplificata dell'immagine iniziale. Per ogni neurone dello strato convoluzionale la rete compie questa procedura e si va quindi a formare per ogni strato un insieme di activation maps di dimensione ridotta. A seguire dopo lo strato convoluzionale, la rete può disporre di uno strato di pooling che ha il compito di ridurre la dimensione dell'output degli strati convoluzionali. In questo passaggio viene usata una funzione di pooling che sintetizza i valori di output dello strato con un indice statistico (max value, media, norma o minimo). Questo ultimo strato permette di ottenere una riduzione tale da poter procedere dopo la successione di strati convoluzionale con un'ultima parte di rete FFNN (*feed forward neural network*) che porterà all'output finale.

2.1 Preparazione dei dati

Le immagini che sono state inserite all'interno dell'algoritmo sono relative alle frequenze dei suoni che ogni traccia emette durante i trenta secondi della durata. Una rete convoluzionale prende in input un tensore che abbia per dimensione la terna:

altezza, larghezza e profondità dell'immagine. I dati sono stati formattati in maniera appropriata per poi essere trasformati in tensori e utilizzati come input nella rete neurale. Si è effettuata quindi una lettura dei file delle immagini, una trasformata delle stesse immagini da JPEG a qualità RGB (griglia di pixels di profondità 3) e una conversione in tensori riscaldando i pixel in un intervallo tra $[0, 1]$. I tensori creati e dati come input all'algoritmo sono stati scelti di dimensione: (288, 432, 3).

2.2 Procedimento

Si è individuata come misura di perdita per il procedimento di ottimizzazione dei pesi, l'entropia incrociata di tipo categoriale, *categorical-cross entropy* sulla base del tipo di variabile target (di tipo categoriale non binaria ma a dieci classi) e per lo stesso motivo l'accuracy come metrica di valutazione.

In una prima fase è stata studiata la struttura esterna della rete, generando dapprima una rete semplice non troppo profonda, andando sempre più nel *deep* aggiungendo e variando i layers per migliorare l'adattamento dell'algoritmo ai dati di training e monitorando man mano l'accuracy sul validation.

Dopo quest'analisi sulla profondità e struttura della rete si è giunti al risultato che la struttura della rete che meglio si è adattata ai dati, pur mantenendo una buona generalizzazione, è composta da tre strati convoluzionali ognuno seguito da altrettanti strati di *max pooling*, legati tutti dalla funzione di attivazione *relu*. In seguito uno strato *flatten* che anticipava gli ultimi due strati *feed forward*. Nell'ultimo strato la rete utilizza una funzione *softmax* necessaria per poter trasformare quanto ottenuto dagli strati precedenti in un output probabilistico di tipo categoriale che potesse dare origine alle etichette della variabile risposta in esame (*genere musicale*). Si è optato infine poi come processo di ottimizzazione il metodo adattivo *Adam* (ovvero *Adaptive momentum estimation*). Andando a inserire diversi tipi di iperparametri i risultati non variavano in maniera ingente (questo a testimonianza della buona robustezza della rete), tuttavia una certa differenza si notava e quindi si è andati a ricercare quei valori degli iperparametri che assicuravano una migliore performance di classificazione nei dati di validation set. E' stato inoltre necessario inserire dei parametri che evitassero il fenomeno

RICONOSCIMENTO DEI GENERI MUSICALI E DELLE LORO CARATTERISTICHE TRAMITE IL DEEP LEARNING

di overfitting della rete perchè infatti senza di essi l'accuracy del training risultava prossima a 1 ma molto distante dal risultato nel validation dove non superava lo 0.4, testimoniando quindi un'ottima capacità di fitting a discapito dell'abilità nel generalizzare. Si sono quindi inseriti uno strato di drop out, dei parametri di learning rate per controllare l'ottimizzazione dei pesi e un parametro di regolarizzazione l2 dei pesi nel penultimo strato *feed forward*. Il termine "Drop-Out" si riferisce all'abbandono di unità (sia nascoste che visibili) in una rete neurale. La procedura fa ignorare i neuroni durante la fase di allenamento in maniera casuale. Più tecnicamente, durante le fasi di addestramento, i singoli nodi vengono eliminati dalla rete con probabilità 1-p o mantenuti con probabilità p, in modo da lasciare una rete ridotta. La probabilità p costituisce un valore da inserire. L'operazione di *tuning* ha quindi coinvolto anche questi parametri, e ha portato al risultato della rete convoluzionale ottima formata da:

Primo strato convoluzionale	32
Secondo strato convoluzionale	64
Terzo strato convoluzionale	128
Dimensione del primo kernel	(4, 4)
Dimensione del secondo kernel	(3, 3)
Dimensione del terzo kernel	(2, 2)
Neuroni nel primo strato convoluzionale	512
Parametro del drop out	(0.4)
Parametro di regolarizzazione l2	0.0001
Parametro di learning rate	0.0001

Come illustrato nella figura 3, l'algoritmo su quaranta epoche ha raggiunto un'ottimo esito di classificazione testimoniato da una accuracy finale sul training di 0.975 e sul validation stabilizzata nelle ultime epoche intorno allo 0.6 con picchi di 0.6375. Le classificazioni effettuate sul test set mostrano un risultato dell'accuracy che si mantiene rilevante: 58% come si evince dalla matrice di confusione mostrata in figura4 che mette in relazione i reali generi di appartenenza delle tracce del test set con le etichette predette.

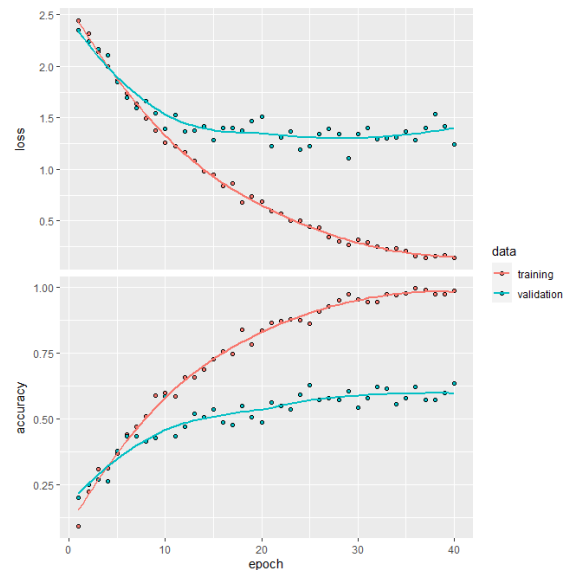


Figura 3: Accuracy e Loss CNN

Come illustrato in figura4 il genere che viene classificato in maniera più corretta è la musical classica con una specificità del 96.67% e una sensibilità del 90% a testimonianza della presenza di una grande differenza tra i suoni di questo genere musicale (più tipici di orchestre) e il periodo storico di maggiore riferimento (prima dell'inizio del novecento) rispetto a tutti gli altri generi musicali. Questa diversità ha portato ad una più semplice classificazione delle tracce relative a questo genere. Per lo stesso motivo, generi come il rock e il blues che sono stati molto presenti durante quasi tutto il ventesimo e ventunesimo secolo, dai quali sono nati altri generi (tra i dieci della variabile) e quindi con caratteristiche acustiche molto simili e confondibili con altri, mostrano delle sensitivity più basse rispettivamente del 30% e 20%.

ACC:0.58	TEORICHE									
PREDETTE	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	2	0	0	0	0	0	0	0	0	0
classical	0	9	1	0	0	0	0	0	0	0
country	1	1	3	2	0	0	0	0	0	3
disco	1	0	4	6	0	0	2	0	0	0
hiphop	0	0	0	1	5	0	0	0	1	0
jazz	0	0	1	0	0	7	0	0	0	0
metal	0	0	0	0	0	0	7	0	0	2
pop	0	0	1	0	1	0	0	8	0	0
reggae	4	0	1	1	4	0	0	2	8	2
rock	2	0	0	0	0	0	1	0	1	3

Figura 4: Matrice Confusione Cnn

3 Feed Forward Neural Network

3.1 I file audio

Come già anticipato nella sezione 1, il dataset contiene mille registrazioni di tracce musicali; per analizzarle e poter cogliere le particolarità che meglio le riescono a contraddistinguere, si è deciso di optare per algoritmi in grado di estrarne le loro caratteristiche. Ognuna di esse ha il merito di riconoscere i segnali audio e di estrarre un valore per ogni traccia che riassume le proprietà rappresentate.

Dal punto di vista matematico, l'audio digitale è una serie temporale ovvero un lungo vettore di numeri reali che rappresentano i livelli di pressione sonora nel tempo. Questa serie temporale può essere visualizzata in diversi domini:

- 1 La semplice serie temporale originale:

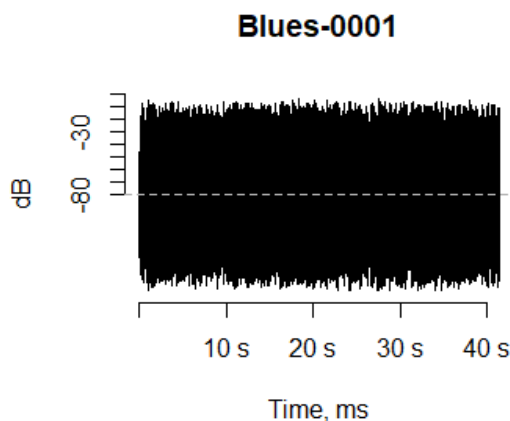


Figura 5: Time Series

- 2 Studio della frequenza. Possiamo usare una Fast Fourier Transform (FFT) per ottenere lo spettro che mostra quali frequenze sono presenti nel segnale. Se vogliamo vedere come questa composizione di frequenze varia nel tempo, produciamo uno spettrogramma come il grafico sopra (definizione di spettrogramma nella sezione precedente). Il metodo più comune per farlo è la trasformata di Fourier di breve durata (STFT). Il risultato è il medesi-

mo che troviamo nelle immagini già a nostra disposizione.

3.2 Analisi Acustica-Estrazione Features

Al centro dell'analisi acustica con il pacchetto da noi utilizzato (soundgen) c'è la trasformata di Fourier di breve durata (STFT): si osserva un breve segmento di suono alla volta (un frame STFT), si analizza il suo spettro utilizzando la trasformata di Fourier veloce (FFT) e quindi ci sposta alla cornice successiva, sovrapponendo le precedenti. Per analizzare un suono con le impostazioni predefinite e tracciarne lo spettrogramma, tutto ciò che serve è la specificazione della sua frequenza di campionamento (il valore utilizzato è di 16000 Hz). Il grafico dell'analisi è qui rappresentato:

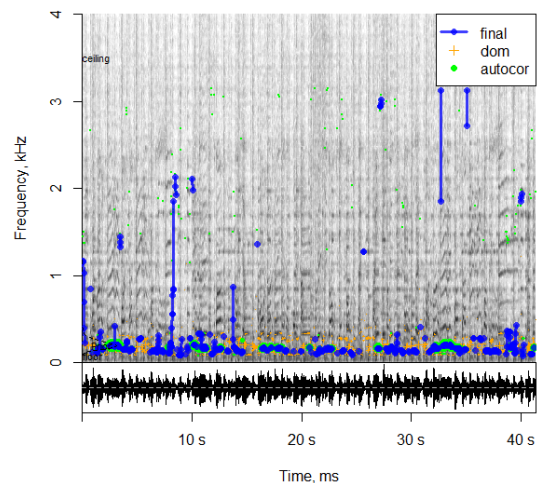


Figura 6: Analisi del tono

Da questa analisi del tono è possibile ricavare 164 features. Si osservi che la traccia *jazz.00054* e *classical.00080* sono state eliminate dall'analisi in quanto il software da noi utilizzato (R-Studio) non è riuscito a riconoscerle correttamente, forse a causa di un loro danneggiamento. Non si elencano tutte in quanto troppe. Alcune di queste sono:

- *time*: tempo della metà di ogni fotogramma (ms)
- *duration*: durata totale (i)

RICONOSCIMENTO DEI GENERI
MUSICALI E DELLE LORO
CARATTERISTICHE TRAMITE IL DEEP LEARNING

- *noSilence*: durata dall'inizio del primo frame STFT non silenzioso alla fine dell'ultimo frame STFT non silenzioso
- *amEnvFreq*, *amEnvPurity*, *amEnvDep*: frequenza (Hz), purezza (dB) e profondità (da 0 a 1) della modulazione di ampiezza stimata da un inviluppo di ampiezza livellato
- *amMsFreq*, *amMsPurity*: uguale a *amEnvFreq* *amEnvPurity*, ma stimato tramite modulation-Spectrum, gli argomenti a cui vengono passati `roughness = list()`
- *amMsFreq*: frequenza di modulazione di ampiezza
- *ampl*: radice quadrata media dell'ampiezza per fotogramma
- *dom*: banda di frequenza dominante più bassa (Hz)
- *entropy*: Entropia Weiner dello spettro del frame corrente. Vicino a 0: tono puro o suono tonale con quasi tutta l'energia in armoniche; vicino a 1: rumore bianco
- *f1 freq*, *f1 width*,...: la frequenza e la larghezza di banda delle prime *n*Formantsformanti per frame STFT
- *flux*: flusso basato sulle caratteristiche, il tasso di variazione delle caratteristiche acustiche come intonazione, HNR, ecc.; senza unità (0 = nessuno)
- *harmEnergy*: la quantità di energia nelle armoniche superiori, ovvero il rapporto tra l'energia spettrale totale superiore a $1,25 \times f_0$ e l'energia spettrale totale inferiore a $1,25 \times f_0$ (dB)
- *harmHeight*: come raggiungono le alte armoniche nello spettro, in base alla migliore ipotesi di intonazione (o ai valori di intonazione forniti manualmente)
- *HNR*: rapporto armoniche/rumore (dB), una misura dell'armonicità. Se $HNR = 0$ dB, c'è tanta energia nelle armoniche quanta nel rumore
- *loudness*: sonorità soggettiva in sone, assumendo un certo livello di pressione sonora (tiene conto dell'energia nelle diverse bande di frequenza e della sensibilità dell'orecchio umano alle diverse frequenze)

Per quanto riguarda il dataset creato ai fini della nostra analisi, considerato che si sarebbe andato ad utilizzare delle reti neurali, è stato importante adottare metodi come standardizzazione e/o nor-

malizzazione. Le variabili in questione hanno range molto diversi il che porterebbe a pesi differenti nell'algoritmo. Si è così deciso di adottare una normalizzazione dei dati. Nessun pre-processing sul dataset è stato effettuato in quanto non necessario per questo tipo di algoritmi.

3.3 Feed Forward Neural Network

Il modello feed forward è la forma più semplice di una rete neurale poiché le informazioni vengono elaborate solo in una direzione sebbene i dati possano passare attraverso più nodi nascosti, si sposta sempre in una direzione e mai all'indietro. Una rete neurale Feed Forward possiede connessioni tra i nodi che non formano un ciclo. L'opposto di questo algoritmo è una rete neurale ricorrente, in cui vengono ciclati determinati percorsi.

Il primo passo necessario è stata la conversione del dataset nel formato necessario affinché si potesse correttamente inserirlo nella rete neurale. È stato così trasformato il tutto in una matrice ed è stata convertita la variabile categoriale target (genere) in una variabile numerica. I dati come già spiegato in precedenza sono stati normalizzati per essere maggiormente confrontabili tra di loro.

Per trovare il miglior tipo di rete è stato effettuato un tuning manuale per i troppi problemi computazionali dovuti ad un'ottimizzazione automatica. È stata così per prima cosa addestrata una rete con un solo hidden layer, variando il numero dei neuroni (128, 256, 512). Un approccio di questo tipo è stato fatto per capire se ci fosse o meno la presenza di overfitting e per valutare il numero di neuroni di partenza.

Successivamente il numero di strati nascosti è stato aumentato progressivamente, al fine di cercare di migliorare le performance nel validation della rete in termini di accuracy. Si è notato fin da subito un grosso problema di overfitting. Per risolvere tale situazione si è optata la tecnica del drop out. Il learning rate è stato tenuto costante pari a 0.0001, con metodo di ottimizzazione "adam" e il numero di epoche pari a 41. Si sono provati differenti parametri per il learning rate che non hanno però portato contributo positivo alla rete neurale.

RICONOSCIMENTO DEI GENERI MUSICALI E DELLE LORO CARATTERISTICHE TRAMITE IL DEEP LEARNING

3.4 Risultati

La rete neurale risultante con le migliori performance sul validation è stata la seguente:

Hidden Layer 1	512 Units	Relu
Layer DropOut	Rate=0.3	
Hidden Layer 2	256 Units	Relu
Layer DropOut	Rate=0.3	
Hidden Layer 3	128 Units	Relu
Layer DropOut	Rate=0.3	
Hidden Layer 4	64 Units	Relu
Layer DropOut	Rate=0.3	
Hidden Layer 5	10 Units	SoftMax

La tecnica del dropout è stata necessaria e fondamentale al fine di prevenire l'overfitting ed è stata utilizzata per ogni strato nascosto.

Le performance della rete neurale in questione possono essere osservate in questo grafico riassuntivo:

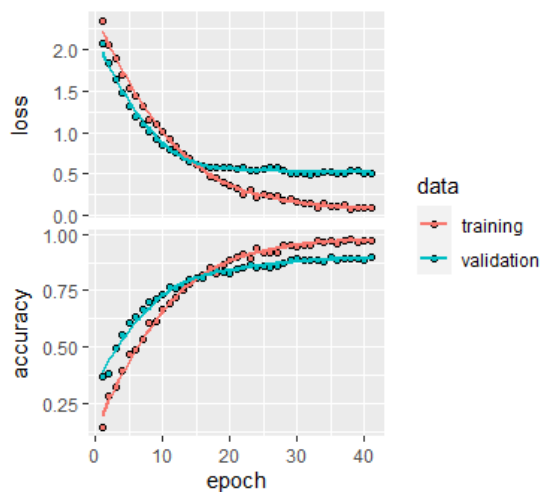


Figura 7: Feed Forward Neural Network

Si può osservare come la rete con i parametri scelti in precedenza raggiunge un ottimo valore di accuracy sul validation pari al 88% circa mentre il problema del overfitting è risolto. Si noti come nelle prime epoche l'accuracy del validation sia più alta rispetto a quella del training. Questo può essere spiegato dalla presenza di istanze molto semplici da classificare nel validation set e da osservazioni complicate da studiare all'interno del training set.

La rete comunque si stabilizza dopo la 25 epoca circa.

Applicando quindi la rete all'interno del test set a nostra disposizione (costituito dal 10% delle osservazioni) si ottiene la seguente matrice di confusione:

ACC:0.9018	TEORICHE									
PRED	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	13	2	1	0	0	0	0	0	0	0
classical	0	7	1	0	0	0	0	0	0	0
country	0	0	12	0	0	0	0	0	0	0
disco	0	0	0	5	0	0	0	0	0	0
hiphop	0	0	0	2	13	0	0	0	0	0
jazz	0	0	0	0	0	9	0	0	0	0
metal	0	0	0	0	0	0	10	0	0	0
pop	0	0	0	0	1	0	1	11	1	1
reggae	0	0	0	0	0	0	0	0	15	1
rock	0	0	0	0	0	0	0	0	0	6

Figura 8: Matrice Confusione FFNN

L'accuracy raggiunta sul test è molto elevata. Per puro aspetto comparativo, si ha testato un random forest sul dataset a nostra disposizione ottenendo un'accuracy sul test pari a 0.615. Questo ci indica quanto il deep learning possa, se ben utilizzato, portare a ottimi risultati anche a confronto con altre tecniche del machine learning.

Conclusioni

I risultati ottenuti nella classificazione hanno evidenziato una precisione massima del 90%, con reti neurali di tipo feed forward applicate alle features estrapolate dalle tracce audio, nel riconoscere un genere musicale.

Il pop, il blues e il rock sono stati i più complicati da etichettare a testimonianza della loro similitudine in termini di melodie, ritmo, strumentazione e periodi storici, a differenza della musica classica e il jazz che hanno effettivamente fatto rilevare una marcata singolarità nelle loro caratteristiche.

Per quanto riguarda le reti convoluzionali si è riusciti ad ottenere un'accuracy del 58%. Le immagini dello spettrogramma sono molto complicate da decifrare e il risultato ottenuto è comunque ottimo. Si noti che considerando una classificazione casuale, essa avrebbe la probabilità di assegnare il genere corretto solo del 10%, l'esito può considerarsi di ottima utilità.; per questo motivo l'utilizzo del deep learning nel seguente ambito si è rivelato di

*RICONOSCIMENTO DEI GENERI
MUSICALI E DELLE LORO
CARATTERISTICHE TRAMITE IL DEEP LEARNING*

grande efficacia e può essere di ottimo aiuto per l'organizzazione di brani musicali sulle piattaforme in streaming (per esempio Spotify) e per applicazioni atte al riconoscimento audio (per esempio Shazam).

Bibliografia e Sitografia

- 1 <https://cran.r-project.org/web/packages/soundecology/vignettes/intro.html>
- 2 https://rdr.io/cran/soundecology/man/multiple_sounds.html
- 3 https://cran.r-project.org/web/packages/soundgen/vignettes/acoustic_analysis.html
- 4 <https://www.sicara.ai/blog/2019-10-31-convolutional-layer-convolution-kernel#:~:text=A%20common%20choice%20is%20to,%3A%203%2C%201%20by%20color>
- 5 <https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/#:~:text=Dropout%20is%20a%20technique%20where,%E2%80%9Cdropped%20out%E2%80%9D%20randomly.>
- 6 <https://elearning.unimib.it/course/view.php?id=38065>