

ECG classification with CNN

Bianchini Linda, Cabras Laura
June 23, 2021

Abstract

Il progetto consiste in una rete neurale convoluzionale per la classificazione degli ECG di 294 pazienti. In particolare per evidenziare quelli con infarto del miocardio (MI), attacco cardiaco che deriva dall'interruzione del flusso sanguigno tissutale causato dall'occlusione delle arterie coronarie a causa di un trombo. I danni che può causare al soggetto in esame possono essere di tipo permanente e causare la morte del tessuto del miocardio. E' possibile ridurre la gravità dei danni se c'è un intervento immediato.

Il proposito della rete è quindi di identificare i MI. Il miocardio è un muscolo esteso perciò si possono osservare vari tipi di infarto in base alla posizione dell'occlusione.

E' stato dunque utilizzato il dataset *PTB diagnostic ECG dataset* che contiene 294 pazienti di cui 52 soggetti sani e 148 pazienti con infarto miocardico, e per ognuno di questi 12 derivazioni. In particolare abbiamo scelto di utilizzare la seconda tra le 12, poiché è comunemente utilizzata nella classificazione dei segnali. Le derivazioni rappresentano i segnali ottenuti ponendo degli elettrodi in punti diversi del torace. Abbiamo quindi considerato il singolo battito nell'identificazione dell'infarto miocardico.

Nella figura sottostante sono sovrapposti due battiti, uno di una persona sana e uno di un'altra con infarto del miocardio.

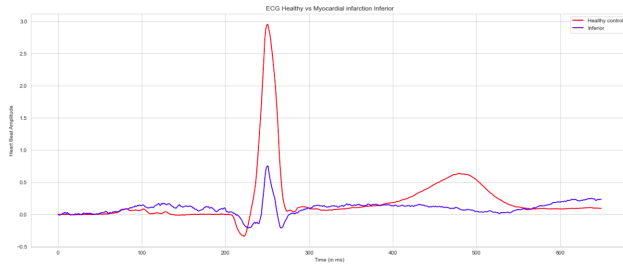


Figure 1: In rosso il battito di una persona sana e in blu il battito di una con infarto miocardico.

1 TASKS

Il progetto è strutturato su più punti:

- Data processing
- Applicazione della rete descritta nel paper di riferimento [?]
- Variazione del learning rate, del numero di layer interni della rete e dei parametri corrispondenti
- Applicazione della regressione L1 ed L2 e valutazione delle grandezze dalla confusion matrix

2 METODI

In questa prima parte del progetto viene illustrato il data preprocessing realizzato sui dati 'PTB diagnostic ECG Database v1.0.0'. In questo dataset vi sono 290 pazienti, numerati da 1 a 294, mancano infatti i pazienti 124, 132, 134 e 161. Tra questi 148 presentano infarto miocardico, 52 sono sani e gli altri 90 presentano altri tipi di anomalie. I dati erano inizialmente in un file `.npz` e salvati con il numero del paziente e il nome del file, entrambi forniti dalle prime due colonne del file `meta.csv`. L'obiettivo è stato isolare i singoli battiti e mettervi i label.

Per quanto riguarda l'isolamento del singolo battito si è utilizzata libreria `neurokit2`, ottimizzata nell'identificazione dei picchi R. Una volta identificato il picco principale si sono presi 250 campionamenti prima del picco e 400 dopo, come suggerito dal paper di riferimento.

La scelta della finestra di campionamenti è dovuta alle zone di interesse nell'individuazione dell'infarto miocardico, ossia il picco R e l'onda T.

Per quanto riguarda invece il posizionamento dei label sono state considerate 12 classi: una degli individui sani e le altre 11 in base alla posizione dell'infarto miocardico. Le classi sono le seguenti: anteriore (A), anteriore laterale (AL), anteriore del setto (AS), anteriore del setto laterale (ASL), inferiore (I), laterale inferiore (IL), posteriore inferiore (IP), laterale posteriore inferiore (IPL), laterale (L), posteriore (P) e laterale posteriore (PL).

Inizialmente si è considerato un singolo file per paziente, per osservare la distribuzione dei pazienti stessi nelle classi, è infatti stato verificato che la diagnosi del singolo paziente non variasse tra due file diversi.

I dati risultavano sbilanciati con una maggioranza di casi sani, perciò si sono considerati anche gli altri file dei pazienti con infarto miocardico, aumentando così il numero di battiti totali (Fig. 11).

Considerando il paper di riferimento a questo punto si è progettata una CNN per il riconoscimento dei segnali, la rete nel nostro caso viene addestrata in particolare con un elettrodo, il secondo, che è uno dei più rappresentativi, ma potenzialmente realizzabile per ciascun segnale delle derivazioni.

Il modello di CNN implementato è composto da 10 strati:

- due strati di Convoluzione 1D
- Maxpooling 1D
- Dropout (20 %)
- due strati di Convoluzione 1D
- Maxpooling 1D

- Flatten
- Dense
- Softmax

In questa prima architettura sono stati mandati in input i dati precedentemente descritti, nonostante non fossero ancora bilanciati tra le classi. Vengono suddivisi in un 70% di training set, 15% di validation set e 15% di test set.

Per migliorare il modello si effettua un data augmentation (Fig. 12) e si aggiunge un altro layer di Dropout al 20% prima del layer Dense.

La rete proposta è molto ampia, con 934796 parametri, siamo nel caso del Deep Double Descent, si è perciò deciso di diminuire i parametri fino al di sotto del numero dei dati (120000 con data augmentation) per analizzare più agevolmente la rete.

La nuova rete diminuita elimina i doppi layer di convoluzione 1D lasciandone solo uno.

Per ottenere un modello che fosse in grado di generalizzare e ottimizzare abbiamo variato il learning rate η del SGD. Per ciascuna fase di addestramento abbiamo utilizzato batch di dimensione 64 e 50 epoche, quindi di numero inferiore rispetto alla dimensione del training set di dati per evitare l'overfit.

Il learning rate regola l'ampiezza della dimensione dello step di modifica dei pesi

$$\omega_{t+1} = \omega_t - \eta \nabla_{\omega} J(\omega) \quad (1)$$

Abbiamo quindi impostato η di 0.01, 0.001, 0.1, 0.005, 0.05 e 0.008. Per ognuno di questi abbiamo graficato l'accuracy, la loss function e la confusion matrix (Fig. 2). Abbiamo poi generato un grafico complessivo dei risultati ottenuti per loss function e accuracy e abbiamo visto che l'ottimizzazione migliore ottenuta è con un learning rate di 0.01, che porta ad una accuratezza maggiore del 90%.

Dopo aver variato il learning rate si è proseguito ottimizzando l'architettura della rete ridotta, variando i parametri corrispondenti al layer di convoluzione centrale. Si è cercato di mantenere il numero di parametri totali attorno a 60000, inferiore al numero dei dati. Confrontando sempre accuracy e loss, il risultato migliore è stato ottenuto con un numero di parametri di 50076.

Abbiamo applicato le tecniche di regolarizzazione L1 (LASSO) ed L2 (RIDGE), che operano sulla funzione costo e permettono di migliorare la rete e diminuire l'overfit. Prevedono l'aggiunta di un fattore, dipendente dai pesi, aggiunto all'espressione della funzione costo.

$$J(\omega) = \frac{1}{2m} [\|\hat{y} - y\|^2 + \lambda \|\omega_j\|] \quad (2)$$

con $\lambda \|\omega_j\|$ dato dalla Regolarizzazione L1.

$$J(\omega) = \frac{1}{2m} [\|\hat{y} - y\|^2 + \lambda \|\omega_j\|^2] \quad (3)$$

con $\lambda \|\omega_j\|^2$ dato dalla Regolarizzazione L2.

Dove λ è un iperparametro chiamato tasso di regolarizzazione. Abbiamo quindi applicato le due regolarizzazioni alla rete di layers diminuita, fissando l'iperparametro λ a 0.01, 0.001, 0.02, 0.005 e abbiamo generato le rispettive confusion matrix, i grafici di accuracy e di loss. Ma si è osservato che sovrapponendo i grafici (Fig. 4-5), i risultati migliori sono dati dalla rete priva di regolarizzazioni sia per L1 che per L2.

In ultima analisi per valutare i modelli utilizzati si sono considerate le grandezze ricavabili dalla confusion matrix: accuracy, precision, recall e F1-score. In particolare vediamo quantità micro, macro e weighted. Le micro corrispondono alle grandezze calcolate per ogni classe senza alcuna media. Le macro calcolano le metriche per ogni classe individualmente e quindi prende la media non ponderata delle misure. Le weighted invece fanno anche la media ponderata. I valori ottenuti per accuracy e per le altre grandezze nel calcolo weighted sono stati riportati in Tab. 1 nell'Appendice.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TN}{TN + FP} \quad (6)$$

$$F1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

3 RISULTATI

Dalle prime reti complete abbiamo ottenuto un overfit come si vede dall'immagine in appendice (Fig. 6), la prima rete non prevede bene le classi con label 8, 9, 10. Abbiamo perciò effettuato data augmentation ottenendo nuovamente un overfit (Fig. 7), per questo motivo si è deciso di aggiungere il layer di Dropout al 20%. L'ultima rete infatti predice bene le classi del test set (Fig 10 Appendice), abbiamo un'alta concentrazione nella diagonale principale.

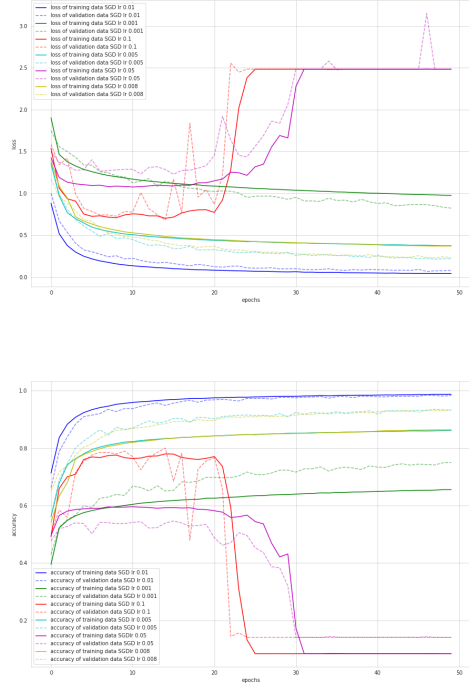


Figure 2

I grafici della Fig. 2 riassumono l'accuracy e la loss delle varie reti in cui abbiamo variato il learning rate. Si vede che la traccia blu risulta quella migliore, con learning rate 0.01. In particolare si nota che la rossa e la viola rispettivamente 0.1 e 0.05 non vadano a convergenza. Si è poi variato il layer centrale di convoluzione, in particolare ottenendo 71724, 64300, 36060, 50076 parametri.

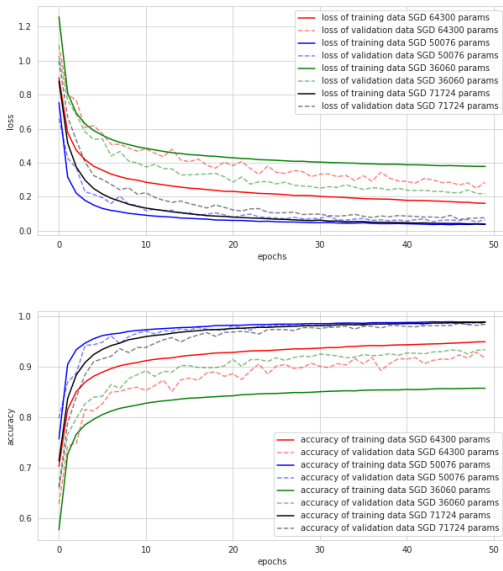


Figure 3

In questi grafici di Fig. 3 riassuntivi si nota che la traccia

nera (71724 parametri) e quella blu (50076 parametri) sono quelle che hanno dato risultati migliori anche confrontando le confusion matrix in appendice (Fig.). Si è dunque deciso di mantenere il modello con 50076 parametri.

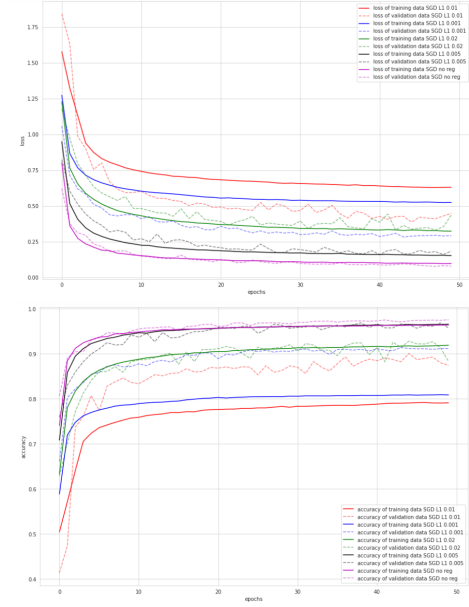


Figure 4

Dai grafici ottenuti per la regolarizzazione L1 (Fig. 4) si osserva che i risultati migliori si ottengono per la traccia viola (senza regolarizzazione), quella nera ($\lambda = 0.005$) e infine per quella verde ($\lambda = 0.02$).

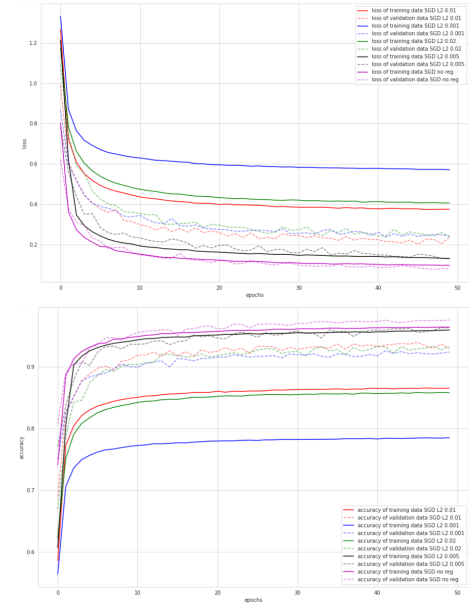


Figure 5

Nella regolarizzazione L2 (Fig. 5) vediamo che i tracciati migliori risultano quelli viola e neri: rispettivamente senza

regressione e con $\lambda = 0.005$.

Infine nella Tab. 1 sono stati riportati i valori di accuracy, precision, recall ed F1-score delle ultime reti effettuate con regolarizzazione. I valori più alti si ottengono per la rete senza regolarizzazione e quella con regolarizzazione L2 e $\lambda = 0.005$.

4 CONCLUSIONI

Per riassumere in questo progetto si è cercato di riprodurre la rete del paper di riferimento e valutarne alcune possibili ottimizzazioni. A questo scopo l'architettura proposta è stata diminuita per renderla più agevole da studiare.

Dopo una prima parte di data preprocessing si sono variate alcune caratteristiche della rete variando il numero di parametri relativi ai layer interni e svolgendo quindi delle ottimizzazioni sia tramite SGD, studiando diversi valori di learning rate, che con regolarizzazioni L1 e L2.

L'obiettivo è stato quello di valutare come la nostra rete cambiasse in funzione delle modifiche appena descritte. Riteniamo che la rete implementata, anche se con meno layer rispetto a quella sviluppata dal paper di riferimento, sia efficiente nell'individuazione degli ECG con MI.

Un possibile sviluppo del progetto potrebbe essere quello di migliorare i dati iniziali attraverso trasformate Wavelet e di Fourier e ricostruire le onde con meno rumore.

REFERENCES

- [1] A.L. Goldberger, L.A. Amaral, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, *Circulation* 101 (23) (2000) e215–e220 <https://www.kaggle.com/openmark/ptb-diagnostic-ecg-database?select=meta.csv>
- [2] Ulas Baran Baloglu, Muhammed Talo, Ozal Yildirim, Ru San Tan, U Rajendra Acharya, *Classification of myocardial infarction with multi-lead ECG signals and deep CNN*, *Pattern Recognition Letters*, Volume 122, 2019, Pages 23-30, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2019.02.016>. (<https://www.sciencedirect.com/science/article/pii/S016786551930056X>)

APPENDICE

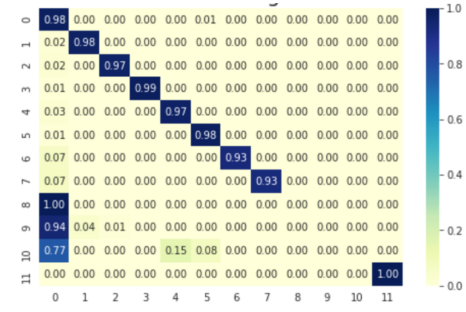


Figure 6: Confusion matrix della rete senza data augmentation, con l'architettura completa descritta dal paper di riferimento

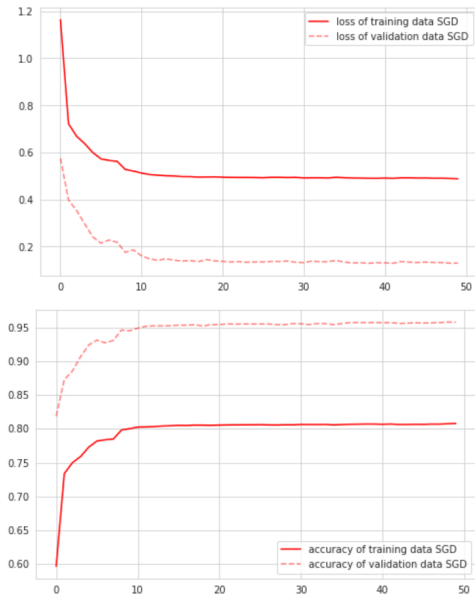


Figure 7: Loss e accuracy della rete con data augmentation e architettura completa. Presenta un evidente overfit.

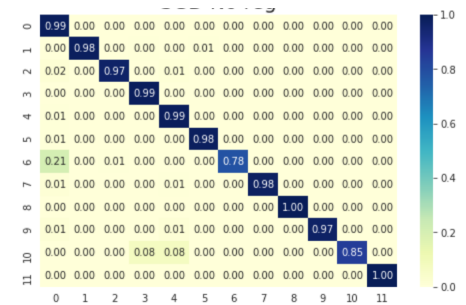


Figure 8: Confusion matrix della rete con architettura completa e dati aumentati, con aggiunta di layer di dropout.

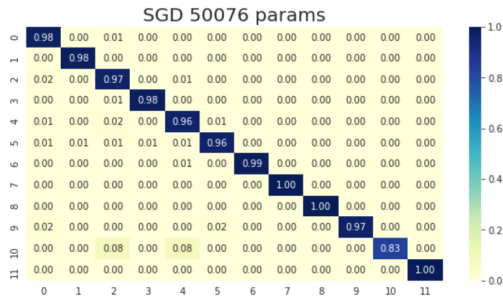


Figure 9: Confusion matrix della rete con architettura diminuita, con 50076 parametri.

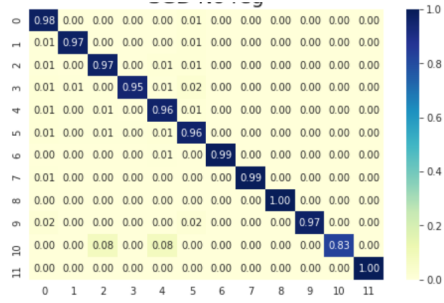


Figure 10: Confusion matrix della rete con architettura diminuita, con 71724 parametri.

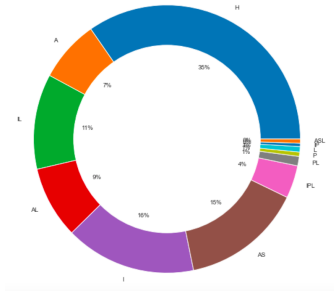


Figure 11: Distribuzione iniziale dei dati nelle 12 classi attraverso un diagramma a torta.



Figure 12: Distribuzione dei dati nelle 12 classi attraverso diagramma a torta dopo data augmentation.

Table 1: Grandezze ricavate dalla confusion matrix

Regolarizzazione	λ	Accuracy	Macro precision	Macro sensibility	Macro F1-score
No regolarizzazione	—	0.97	0.97	0.97	0.97
L1	0.01	0.87	0.88	0.87	0.87
L1	0.001	0.90	0.88	0.90	0.89
L1	0.02	0.88	0.89	0.88	0.88
L1	0.005	0.95	0.95	0.95	0.95
L2	0.01	0.92	0.92	0.92	0.91
L2	0.001	0.92	0.91	0.92	0.91
L2	0.02	0.93	0.93	0.93	0.93
L2	0.005	0.96	0.96	0.96	0.96