

# Report Challenge 3

Elisa Coceani

19 maggio 2024

## 1 Introduzione

L'obiettivo della challenge è valutare diversi modelli sul dataset KMNIST, contenente immagini da 28\*28 pixel riguardanti caratteri Kanji giapponesi scritti a mano. I modelli devono identificare correttamente la tipologia di carattere tra le 10 possibili.

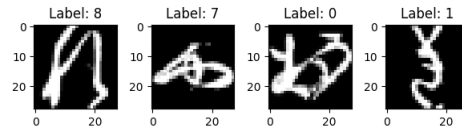


Figura 1: Esempio di immagini del dataset

## 2 Hidden layers, funzioni di attivazione, ottimizzatori e learning rates

Sono stati allenati tre modelli di neural network fully connected con 1, 2 e 3 hidden layer per valutare quale sia il numero di layer più opportuno per svolgere il task assegnato. La stessa operazione è stata svolta con modelli convolutivi.

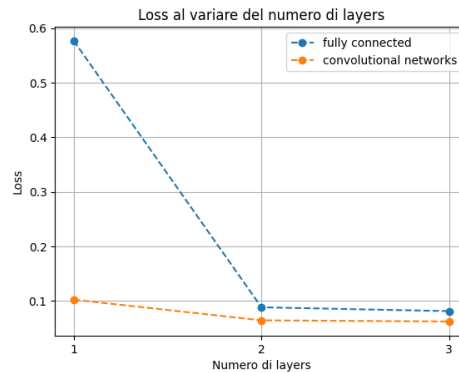
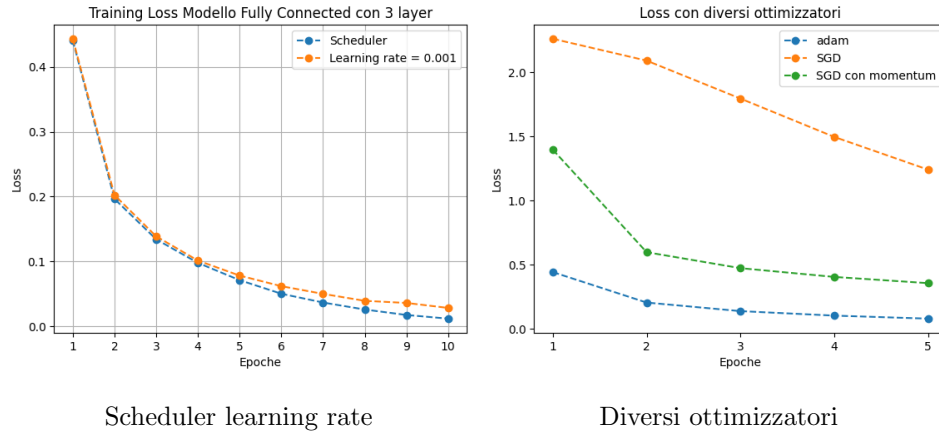


Figura 2: Loss al variare dei layers

Come si vede dal plot della loss durante il training, un modello fully connected con un solo hidden layer è troppo semplice per classificare in modo soddisfacente le immagini. I modelli convolutivi sono i più adatti a questa operazione perché, sfruttando i layer convolutivi, riescono a comprendere meglio la struttura dei dati e classificare correttamente più immagini.

Per i modelli con tre hidden layer la funzione di attivazione Relu risulta quella che massimizza l'accuracy. Per quanto riguarda il learning rate, il più appropriato da utilizzare è 0.001, ma implementando lo scheduler Exponential che diminuisce il learning rate iniziale (0.001) dopo ogni epoca di un fattore  $\gamma = 0.9$ , c'è un leggero miglioramento nella loss:



Fino ad ora è stato utilizzato l'ottimizzatore Adam; provando ad allenare un modello con stochastic gradient descent, stochastic gradient descent con momentum e ottimizzazione Adam, come mostrato nell'immagine sopra, si vede che quest'ultimo resta l'ottimizzatore conveniente da usare per minimizzare la loss.

### 3 Train vs Test

Per il modello convolutivo con due hidden layer è stato plottato la loss e l'accuracy durante il training e durante il test:

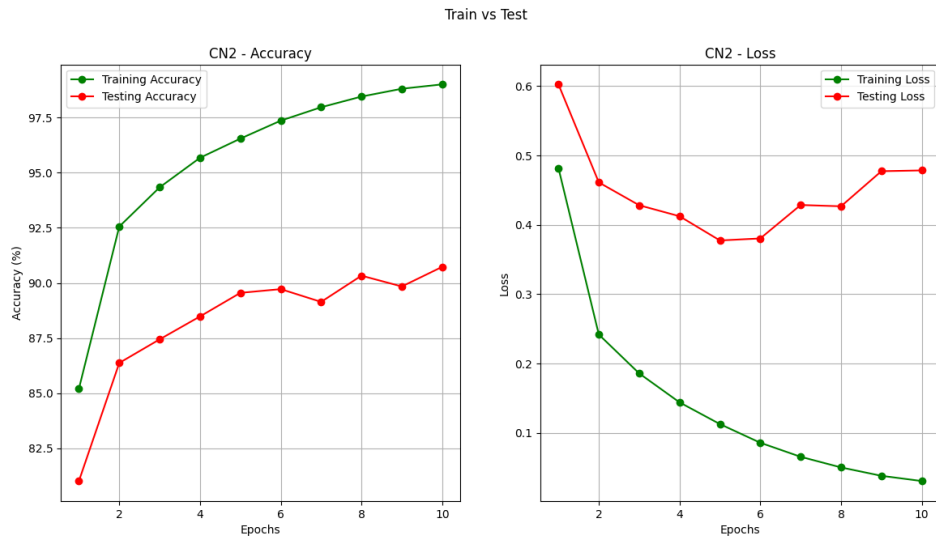


Figura 3: Training vs test accuracy e loss

Il risultato è complessivamente buono, ma nel test set il valore dell'accuracy è più bassa di quanto visto nel train set di circa 8 punti percentuali. Una situazione simile si ripresenta nel plot della loss, anche se con un distacco minore. Questo è dovuto dalla presenza di overfitting: una soluzione potrebbe essere ridurre la complessità del modello introducendo una regolarizzazione. Inoltre con l'aumentare delle epoche di training non ci sono miglioramenti nella loss test, ciò indica che, per prevenire overfitting, potremmo fermare il training alla settima epoca.