



UNIVERSITÀ DEGLI STUDI DI SALERNO

MASTER DEGREE IN BIONFORMATICS AND DIGITAL HEALTH

NATURAL COMPUTATION

MultiLayer Perceptron

Authors:

Alessandro MONTEFUSCO

Vincenzo FERRARA

Supervisor:

Prof. Angelo MARCELLI

second semester 2019/2020

Introduzione

Per l'attività di laboratorio prevista, i parametri del MLP sono stati scelti secondo le specifiche assegnate, variando:

- Learning Rate;
- Numero di Neuroni;
- La dimensione del Training Set e Validation Set.

Learning rate

Il learning rate è stato variato con passo 0.15 (0.30, 0.45, 0.6), in quanto non vi erano variazioni significative tra le diverse configurazioni.

Numero di neuroni

La rete è composta da un livello di input, un livello hidden e un livello di output.

Il primo esperimento è stato effettuato fissando il numero di neuroni come la media tra neuroni di input e di output (32 neuroni).

Il secondo esperimento è stato effettuato fissando il numero di neuroni pari alle features che descrivono un singolo task (22 neuroni).

Training Set e Validation Set

Le configurazioni scelte per la suddivisione del dataset "*alltask_features*" sono le seguenti:

Dataset	TS	VS	Test		TS	VS	Test
DS3	30	30	40		41	41	54
DS7	35	25	40		48	34	54

Come si può notare, il Test Set rimane fisso a 40%, mentre si hanno delle variazioni tra Training Set e Validation Set. In tal modo, sarà possibile effettuare delle considerazioni sulle prestazioni dell'algoritmo, dipendenti dal numero di campioni di Training.

Le performance ottenute per ogni configurazione sono le medie di 10 sperimentazioni con la stessa configurazione ma al variare del seed.

Considerazioni sulle performance

Variando il learning rate è stato notato che la differenza in termini di *accuracy*, *specificity* e *sensitivity* è minima. Per tale ragione è stato deciso di testare le diverse configurazioni della rete variando il learning rate nel range [0.3; 0.6] con passo 0.15.

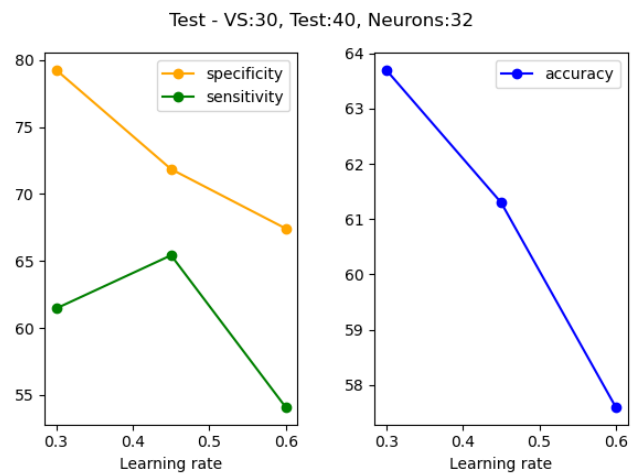
Configurazione: TS 35% - VS 30% - Test 40%

1) Numero neuroni 32

Tale configurazione prevede una rete MLP con un numero di neuroni hidden pari a 32. Quindi, fissata la suddetta configurazione si è notato come variano le prestazioni al variare del learning rate.

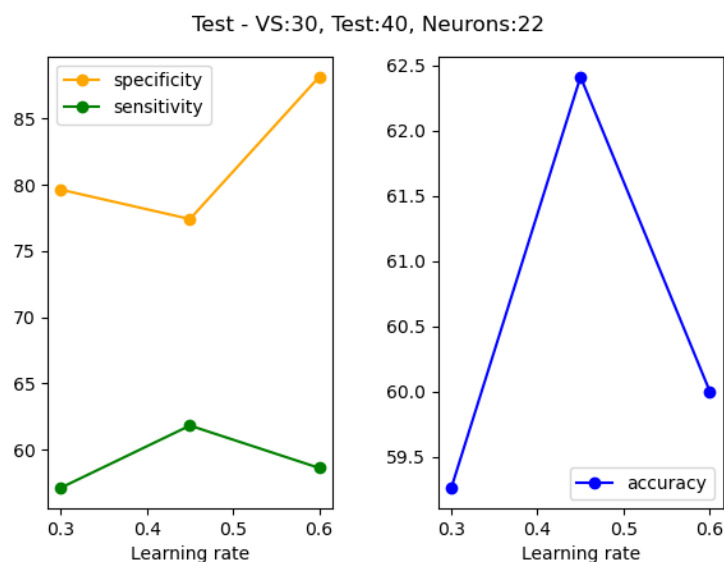
Al crescere del learning rate la percentuale di classificazioni corrette della rete diminuisce di alcuni punti percentuali (circa il 6%). Stesse considerazioni possono essere fatte per la specificity: si può notare che al crescere del learning rate si ha una perdita di più di 10 punti percentuali. Per la sensitivity, invece, si nota un picco con un learning rate pari a 0.45, per poi decrescere all'aumentare dello stesso.

In conclusione, per un Test Set pari al 40% dei campioni del dataset, si ha un modello performante con un valore del learning rate nel range [0.3; 0.45].



2) Numero neuroni 22

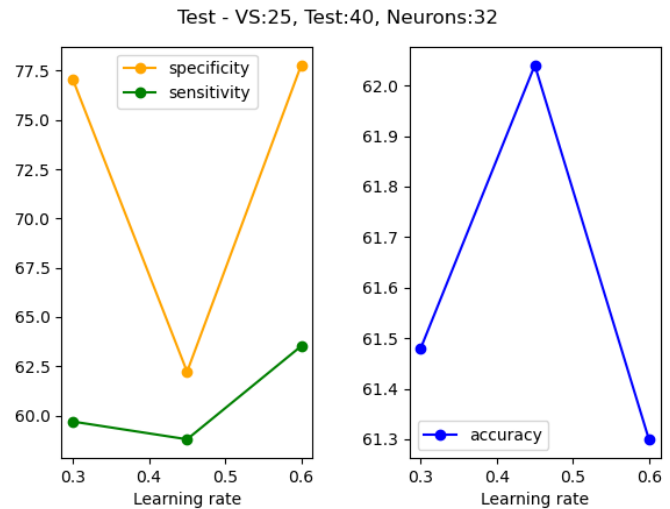
Confrontando tale modello con quello precedente, si può notare che le performance sono mediamente migliori. In particolare, si può affermare che il modello abbia le prestazioni migliori con un learning rate pari a 0.45: l'accuracy è pressoché invariata; la sensitivity è diminuita di pochi punti percentuali; la specificity è nettamente migliorata (+6%). La rete ha un maggior tasso di riconoscimento di TP e un calo sul riconoscimento di TN, ma l'accuracy è circa la stessa.



Configurazione: TS 30% - VS 25% - Test 40%

1) Numero neuroni 32

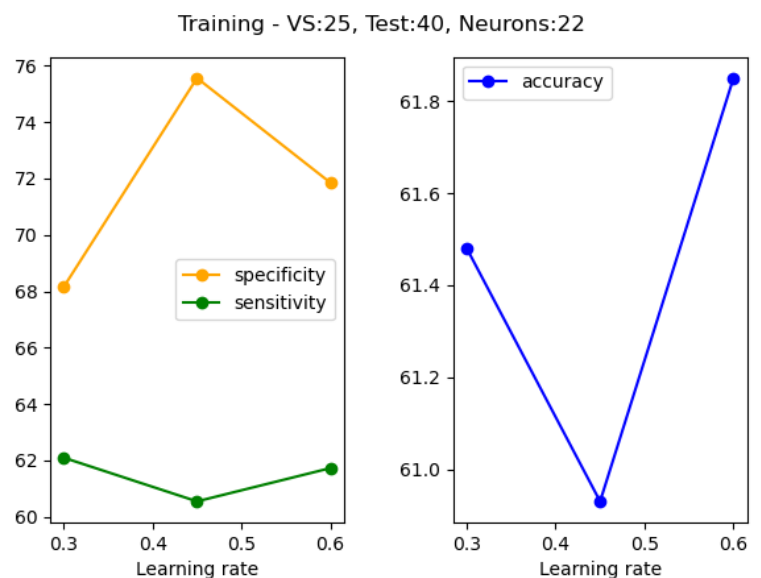
Tale configurazione prevede una rete MLP con un numero di neuroni hidden pari a 32. Quindi, fissata la suddetta configurazione si è notato come variano le prestazioni al variare del learning rate. Al crescere del learning rate la percentuale di classificazioni corrette della rete ha un picco con un learning rate di 0.45 per poi diminuire drasticamente al crescere dello stesso. La specificity ha un andamento opposto: ha un picco negativo con learning rate pari a 0.45 per poi risalire. Infine, per la sensitivity si può notare che al crescere del learning rate si ha una crescita di 2.5 punti percentuali, tranne una leggera flessione con learning rate pari a 0.45.



In conclusione, per un Test Set pari al 40% dei campioni del dataset, si ha un modello performante con un valore del learning rate nel range [0.5; 0.6], poiché l'accuracy ha una variazione minima, mentre si ha un aumento nel riconoscimento di TP e TN.

2) Numero neuroni 22

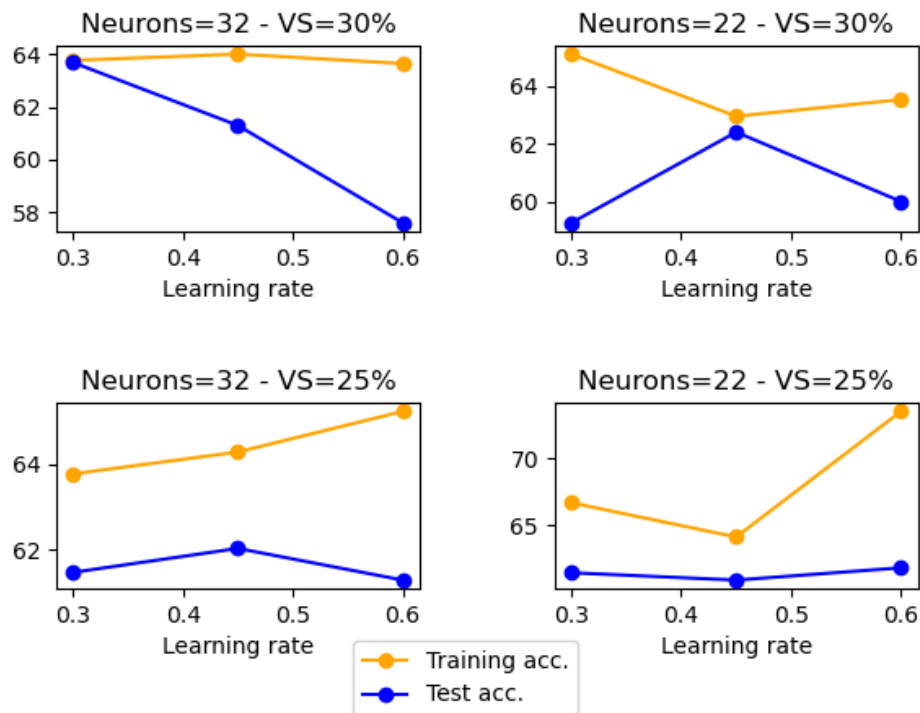
Confrontando tale modello con quello precedente, si può notare che le performance sono mediamente peggiori. In particolare, si può affermare che il modello abbia le prestazioni migliori con un learning rate pari a 0.5: l'accuracy e la sensitivity sono pressoché invariate; la specificity è nettamente migliorata (circa 7%). La rete ha un maggior tasso di riconoscimento di TP, ma l'accuracy è circa la stessa.



Considerazioni finali

Il modello è stato valutato sul Training Set usando una 5-fold cross validation, per poi valutare le performance su nuovi dati forniti alla rete appartenenti al Test Set.

Dagli esperimenti compiuti sulle configurazioni scelte si ottengono i seguenti grafici:



Con la stessa dimensione del Test Set e variando di un 5% la dimensione del Validation Set, si può osservare che le performance della rete sono migliori quando si hanno a disposizione più campioni di Training. Per quanto riguarda il numero di neuroni del livello hidden, si può affermare che un numero inferiore di neuroni fa emergere un modello con performance migliori. Quindi, si può ipotizzare di utilizzare un MLP con un numero di neuroni hidden pari a 22, in quanto le features estratte sono 22 ma ripetute per 3 task differenti: avrò un neurone hidden per ogni feature di ogni task. Inoltre, si è osservato che le performance migliori si ottengono con un learning rate basso, anche se vi è una forte dipendenza dal numero dei campioni di Training.