

UNIVERSITÀ DEGLI STUDI DI  
MILANO-BICOCCA

ADVANCED MACHINE LEARNING

---

## Assignment 5

---

*Autore:*

Federico Manenti - 790032 - f.manenti3@campus.unimib.it

9 dicembre 2019



# Indice

1	Task e Dataset	1
2	Step 1	1
3	Step 2	3

## 1 Task e Dataset

Il task assegnato consiste nell'ottimizzazione di diversi iperparametri di una rete neurale per classificazione binaria. La metrica da utilizzare e migliorare è l'*accuracy* intesa come media calcolata in *ten fold cross validation*.

Il **dataset** utilizzato si compone di 9 features diverse e la variabile target per 100 osservazioni.

## 2 Step 1

Per il primo step la NN utilizzata è formata da due strati nascosti di rispettivamente 4 e 2 neuroni, gli unici iperparametri da ottimizzare sono:

- **Learning Rate:** tra 0.01 e 0.1
- **Momentum:** tra 0.1 e 0.9

In questo primo esperimento si studia il confronto tra due processi *SMBO* con funzioni di acquisizione diverse, Grid Search e Random Search. Per *SMBO* si utilizzeranno 5 configurazioni iniziali casuali (uguali per le due funzioni di acquisizione) e poi 20 iterazioni con le due diverse funzioni, per Grid e Random search si useranno 25 iterazioni. Il modello surrogato scelto per *SMBO* è il *GP* in quanto gli iperparametri da ottimizzare hanno un range continuo, le due funzioni di acquisizione sono *Expected Improvement* e *Probability Improvement*.

Il grafico del *best seen* nel primo esperimento è il seguente:

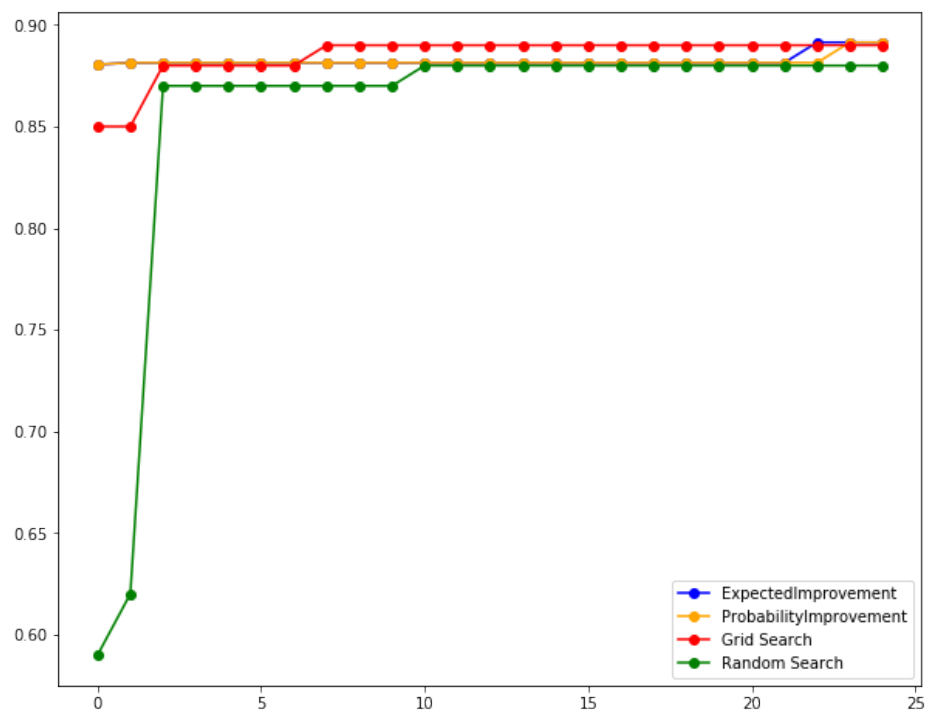


Figura 1: Best seen step 1

Più precisamente i risultati sono:

- **SMBO con expected improvement**
  - Accuracy: 0.90
  - Learning rate: 0.66
  - Momentum: 0.05
- **SMBO con probability improvement**
  - Accuracy: 0.89
  - Learning rate: 0.66
  - Momentum: 0.05
- **Grid search**

- Accuracy: 0.89
- Learning rate: 0.03
- Momentum:0.5

- **Random search**

- Accuracy: 0.88
- Learning rate: 0.20
- Momentum:0.02

Come si può vedere dal grafico e dai risultati riportati i quattro metodi raggiungono valori simili di accuracy, ciò può essere dovuto al fatto che il task di classificazione affrontato è molto semplice e alla scarsità dei dati. Un altro possibile motivo di questi risultati è il forte sbilanciamento della variabile target nel dataset e il non preprocessing che porta la rete neurale a favorire la *Zero Rule*.

### 3 Step 2

Per il secondo esperimento si richiede invece di ottimizzare oltre ai due precedenti iperparametri anche il numero di neuroni (tra 1 e 5) per entrambi i layer nascosti, utilizzando però solo *SMBO* con due funzioni di acquisizione diverse. In questo esperimento il budget è diviso in 10 configurazioni casuali iniziali uguali per le due funzioni e ulteriori 100 iterazioni per ognuna delle funzioni (le stesse del precedente step).

A differenza dello scorso esperimento il modello surrogato utilizzato è il *Random Forest* perché il numero di neuroni per layer è un iperparametro intero e non continuo.

Il grafico del *best seen* in questo caso è:

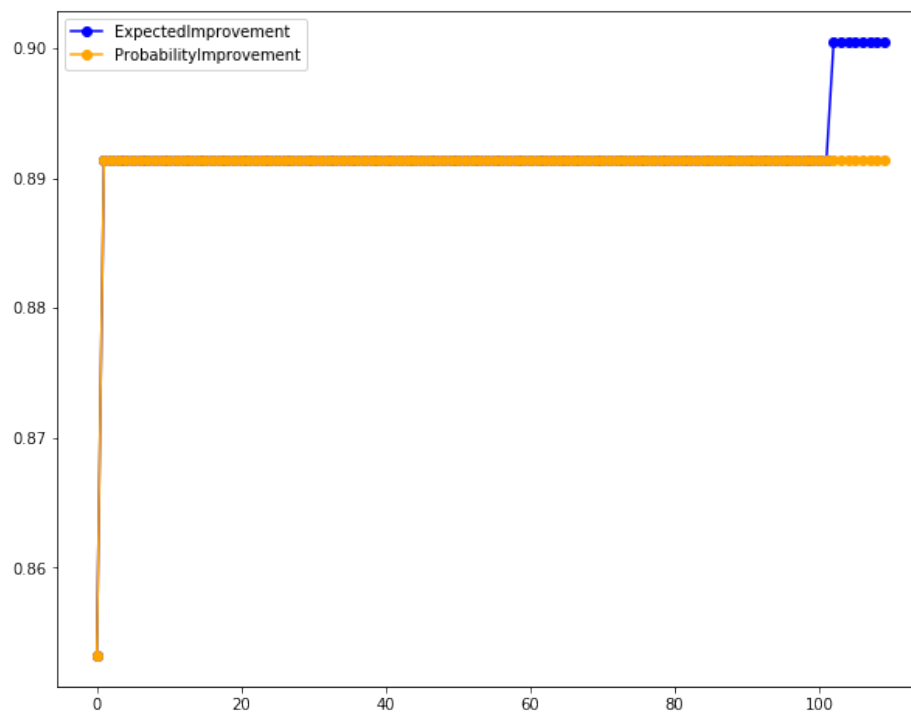


Figura 2: Best seen step 2

I risultati più precisi sono:

- **SMBO con expected improvement**
  - Accuracy: 0.90
  - Learning rate: 0.66
  - Momentum: 0.05
  - Neuroni primo layer: 2
  - Neuroni secondo layer: 2
- **SMBO con probability improvement**
  - Accuracy: 0.89
  - Learning rate: 0.66

- Momentum: 0.05
- Neuroni primo layer: 2
- Neuroni secondo layer: 2

Anche qui i risultati sono praticamente identici e il motivo potrebbe essere ancora lo stesso: dataset piccolo e sbilanciato e task molto semplice.