

Uso del Programma

12 Febbraio 2018

1. Come primo passo si deve scaricare il file che contiene il database Fashion MNIST e estrarre il file zip
N.B.: visto che successivamente si dovrà utilizzare il percorso di alcuni file all'interno del pacchetto suddetto é consigliabile utilizzare percorsi che non contengono spazi separatori o caratteri speciali.
2. Successivamente si dovrà impostare il giusto path dei file da cui verranno estratti il training set e il test set nei file `naive_bayes.py` e `random_forest.py`, nella parte iniziale in cui si fa il loading del dataset nella funzione `load_mnist()`.
3. A seconda di quale classificatore si vuole analizzare si possono settare i parametri d'ingresso della funzione `BernoulliNB()`, presa in analisi da SK Learn. In particolare, a seconda del device che si utilizza, é possibile utilizzare la funzione di parallel computing impostando `n_jobs=x` da un valore di default di 1 a un massimo che corrisponde al massimo numero di CPUs disponibili.
4. Per quanto riguarda il file `random_forest.py` si utilizza per tracciare la learning curve di tale classificatore. L'algoritmo é reso disponibile SK Learn attraverso il metodo `RandomForestClassifier()`. Nel file si può decidere la versione di Random Forest che si desidera (senza o con l'operazione di pruning) e si possono settare altri parametri delle funzioni che sono disponibili sul sito SK Learn, tra i quali la massima profondità degli alberi di decisione e il `n_jobs`.
5. É, inoltre, possibile per ogni funzione di classificazione, per cui si desidera tracciare il grafico, passare come parametro della funzione `plot_curve` un booleano `error`. Questo se ha valore `True`, mostrerà, al termine della procedura di apprendimento, il grafico con sull'asse delle ascisse il numero di esempi di train e sull'asse delle ordinate il tasso di errore. Se si imposta il valore di `error` a `False`, allora il grafico risultante corrisponderà al tasso di successo nell'apprendimento al variare di esempi di train.