

Corso di Fondamenti e Applicazioni del Machine Learning

Progetto da Sviluppare

Con riferimento ai file:

- manuale.csv
- training.csv

eseguire i seguenti task:

1. definire manualmente più classificatori sul file "manuale.csv", adottando almeno due dei modelli illustrati a lezione, e valutando le prestazioni da loro ottenute sullo stesso file "manuale.csv"; dopo aver illustrato i passi per adattare i modelli ai dati, implementare i classificatori in Python, utilizzando eventualmente delle API;
2. verificare che il dataset "training.csv" non contenga osservazioni palesemente errate ed effettuare l'analisi esplorativa del dataset rappresentando i risultati anche in forma grafica (boxplot e/o pairplot e matrice di correlazione);
3. considerando i classificatori progettati ed implementati in Python al punto 1), valutare le performance ottenute da ognuno di essi sul file "training.csv" o su qualche suo sottoinsieme, cercando di ottimizzare le prestazioni dei classificatori;
4. con riferimento al file "training.csv", addestrare tramite Scikit-Learn più classificatori, separando opportunamente i campioni nel training set e nel test set, con l'obiettivo di massimizzare le prestazioni sul test set. Alla fine della fase di addestramento, selezionare il classificatore ritenuto più performante. In sede d'esame sarà fornito un altro file, denominato "real_settings.csv", mirato a testare le prestazioni di tale classificatore.

Per poter sostenere l'esame, consegnare l'intera documentazione prodotta, che dovrà contenere la descrizione dettagliata di tutti gli step effettuati con le relative motivazioni, nonché tutti i sorgenti del codice implementato entro 48 ore dalla data dell'appello scelto.

Pur trattandosi di un lavoro di gruppo, ciascun membro deve comunque essere in grado di spiegare in dettaglio l'intera documentazione prodotta, incluso il codice.