

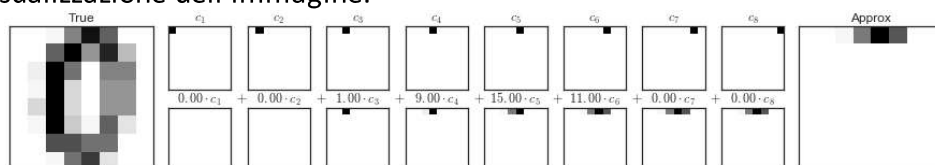
Progetto E

Dato il dataset che contiene le immagini di numeri scritti a mano (8x8=64 pixel), inseriti come righe del file csv. L'ultima colonna di ogni riga rappresenta il target. Il dataset può essere caricato anche da skleran con l'istruzione `datasets.load_digits()`

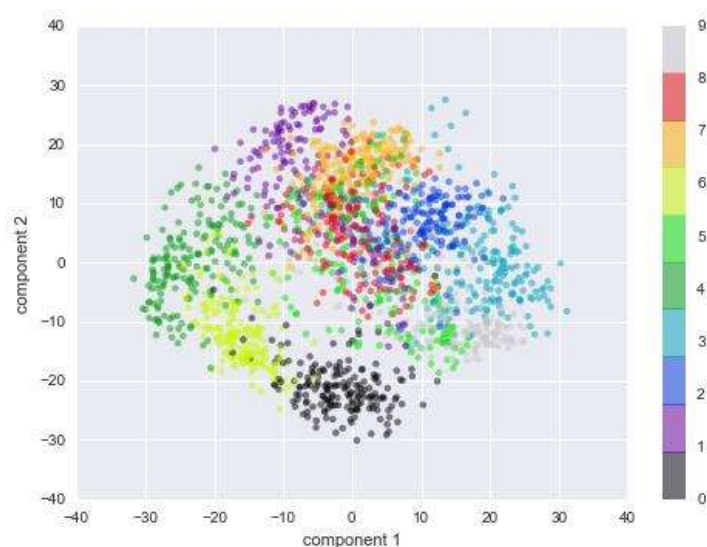
data, {ndarray, dataframe} of (1797, 64) shape: It is a data matrix.

target, {ndarray, series} of (1797,) shape: The target regarding classification.

(<https://www.educative.io/answers/what-is-datasetsloadaddigits-in-sklearn>) l'esempio mostra anche la visualizzazione dell'immagine.



Realizzare un progetto di data analysis che prevede l'utilizzo della PCA per proiettare le 64 dimensioni sulle prime due componenti principali (questi punti sono la proiezione di ciascun punto dati lungo le direzioni con la varianza maggiore). Riprodurre la seguente immagine.



Classificare i punti della proiezione sulle prime due componenti principali tramite metodo SVM (Support Vector Machine). Utilizzare l'80% dei dati per la creazione del modello (training) e il restante 20% per fare il test. Produrre una matrice di confusione che mostri l'accuracy della classificazione.

Dataset:

digits.csv

Output progetto:

- Relazione di progetto (circa 10 pagine) con la descrizione dell'analisi progettuale. Questa deve includere la descrizione delle features, l'intera pipeline di analysis, e risultati ottenuti. Il codice deve essere consegnato su file separati, la relazione può comunque riportare eventuali funzioni di importanza e le relative descrizioni.
- Codice sorgente del progetto.

Modalità di consegna dei progetti:

- Invio e-mail all'indirizzo domenico.garlisi@unipa.it, si suggerisce di specificare nell'oggetto NOME_GRUPPO-ID-PROGETTO, es: MARIO-ROSSI-1 allegando:

o PDF della relazione

o ZIP file contenente i codici sorgente in python