

All'interno di questa cartella sono presenti:

- 2 file Python:
  - `setgenerator.py`: genera il training set e il test set (formattati come richiesto; gli interi sono stati generati casualmente (da 1 a 100), compreso l'output del training set) con una dimensione fissa di ogni entry pari a  $n=10$ ;
  - `predict.py`: effettua l'effettiva predizione per il test set.
- 3 file txt: `trainingset.txt`, `testset.txt`, `output.txt`

All'interno di `predict.py` vengono letti i file `.txt` e vengono salvati in delle variabili (liste di liste). Per ogni entry del test set vengono selezionati i  $k$  elementi del training set meno distanti.

Sia  $X = [x_1, x_2, \dots, x_n]$  un elemento del training set (senza output) e  $T = [t_1, t_2, \dots, t_n]$  un elemento del test set; la nozione di distanza è stata considerata come segue:

$$d(X, T) = \sum_{i=1}^n |x_i - t_i|$$

Una volta selezionati i  $k$  elementi del training set con distanza minima rispetto alla entry del test set, viene creato un nuovo elemento dato dalla media dei  $k$  elementi (**comprendendo l'output**) arrotondata all'intero inferiore. Per esempio (con  $k=3$ ):

$$X^1 = [61, 93, 76, 73, 94, 15, 99, 28, 7, 86, 23]$$

$$X^2 = [27, 52, 60, 94, 11, 16, 90, 41, 8, 83, 77]$$

$$X^3 = [51, 93, 45, 97, 76, 10, 73, 23, 24, 92, 23]$$

↓

$$\bar{X} = [46, 79, 60, 88, 60, 13, 87, 30, 13, 87, 41]$$

A questo punto è stata fatta una proporzione:

$$\bar{x}_{n+1} : \sum_{i=1}^n \bar{x}_i = y : \sum_{i=1}^n t_i$$

↓

$$y = \bar{x}_{n+1} \frac{\sum_{i=1}^n \bar{x}_i}{\sum_{i=1}^n t_i}$$

↓ *Arrotondata all' intero inferiore*

$$y = \lfloor \bar{x}_{n+1} \frac{\sum_{i=1}^n \bar{x}_i}{\sum_{i=1}^n t_i} \rfloor$$

Dove:

- $\bar{x}_{n+1}$ : è l'output dell'elemento ottenuto dalla media dei  $k$  elementi più vicini a  $T$  del training set;
- $\bar{x}_i$ :  $i$ -esima feature dell'elemento ottenuto dalla media dei  $k$  elementi più vicini a  $T$  del training set;
- $t_i$ :  $i$ -esima feature della entry  $T$  del test set;
- $y$ : output per la entry  $T$  del test set.

A questo punto viene salvato il file di output in `output.txt`.