

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING
BIG DATA ANALYTICS
A.A. 2019/2020

Challenge del 18/10/2019
(2 ore e 30 minuti)

Si vuole stabilire il tipo di un dolcetto sulla base di alcune feature. Si leggano i dati dal file `candy.csv`.

Le prime 10 colonne nel file corrispondono alle seguenti caratteristiche:

`caramel`, `peanutyalmondy`, `nougat`, `crispedricewafer`, `hard`, `bar`, `pluribus`,
`sugarpercent`, `pricepercent`, `winpercent`,

L'undicesima colonna del dataset corrisponde all'output in cui il valore 1 indica che si tratta di un dolcetto al cioccolato, mentre il valore 0 indica che non si tratta di un dolcetto al cioccolato.

1. Si normalizzino i dati utilizzando:

- la normalizzazione min-max fra 0 e 5;
- la normalizzazione Z-score.

2. Utilizzando il dataset normalizzato con Z-score, si realizzi a mano un modello di regressione logistica con regolarizzazione L2 e discesa del gradiente (ad unico batch).

3. Al termine dell'addestramento del modello, indicare il valore predetto per la seguente tupla:
(0, 0, 0, 1, 0, 0, 1, 0.87199998, 0.84799999, 49.524113).

4. Si considerino ora anche le ultime due colonne del dataset per i valori di output. In questo modo, le ultime tre colonne avranno valori:

- (1, 0, 0) se si tratta di un dolcetto al cioccolato;
- (0, 1, 0) se si tratta di un dolcetto alla frutta;
- (0, 0, 1) se si tratta di altro tipo di dolcetto.

Si implementi un modello di regressione logistica multiclasse che distingua il tipo di dolcetto in base alle sue caratteristiche.

5. Al termine dell'addestramento del modello, indicare la classe predetta per il dolcetto avente le seguenti caratteristiche:

(1, 0, 0, 0, 1, 0, 0, 0.186, 0.26699999, 41.904308).

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING
BIG DATA ANALYTICS
A.A. 2019/2020

Challenge del 18/10/2019
(2 ore e 30 minuti)

We want to evaluate if a candy is a chocolate one based on some features. Read the data from file `candy.csv`

The first 10 columns correspond to the following characteristics:

caramel, peanutyalmondy, nougat, crispedricewafer, hard, bar, pluribus, sugarpercent, pricepercent, winpercent

while the 11th column corresponds to the output where a value equal to 1 means the candy is a chocolate one.

1. Before the training phase select and apply a normalization to data:
 - min-max normalization between 0 to 5;
 - Z-score normalization
2. Using normalized dataset with Z-score, you have to build a logistic regression model with L2 regularization and full batch gradient descent.
3. After the training phase, compute the predicted value for the following tuple:
(0, 0, 0, 1, 0, 0, 1, 0.87199998, 0.84799999, 49.524113).
4. We want to implement a multi-class classifier where classes are represented in the last three columns. In this way, the last three columns are:
 - (1, 0, 0) if it is a chocolate;
 - (0, 1, 0) if it is fruity;
 - (0, 0, 1) if it is other kind of candy.
5. After the training phase, compute the predicted value for the following tuple:
(1, 0, 0, 0, 1, 0, 0, 0.186, 0.26699999, 41.904308).