

Progetto AMPL

A.A. 2017/2018

1 Materiale

Il dataset è stato creato a partire da due diverse tipologie di dati: un questionario e dei dati oggettivi provenienti da alcuni report di diverse scuole superiori. Complessivamente il dataset è composto da 649 istanze e 33 features.

Gli attributi provenienti dai report sono 4 e consistono nel numero di assenze fatte dallo studente, dal voto ottenuto al termine del primo livello di studi (G1), del secondo livello (G2) e del terzo (G3).

Le features provenienti dal questionario sono in numero decisamente superiore (29) e risultano essere sia di tipo numerico sia nominale.

L'elenco e la descrizione di tutte le features risulta disponibile nella tabella 1.

2 Obiettivo

Utilizzando i dati disponibili descritti in 1 dovrà essere risolto un problema di regressione utilizzando le reti neurali. In particolare l'output del modello dovrà essere, per ogni istanza del training/test, la valutazione al termine del terzo ciclo di studi (G3): dovrà dunque essere un valore numerico compreso tra 0 e 20.

3 Valutazione

Il dataset fornito corrisponde a circa il 70% del dataset originale, il restante 30% verrà utilizzato come test set per la valutazione finale del progetto.

Tabella 1: Elenco e descrizione degli attributi presenti nel dataset

Attributo	Descrizione	Dominio	Range
Sex	Sesso	Binario	"F", "M"
Age	Età	Numerico	15-22
School	Nome Scuola	Binario	"GP", "MS"
Address	Indirizzo di casa	Binario	urbano ("U"), rurale ("R")
Pstatus	Domicilio con i genitori	Binario	Si ("T"), No ("A")
Medu	Livello educazione della madre	Numerico	0-4 ¹
Mjob	Lavoro della madre	Nominale	Nota ²
Fedu	Livello educazione del padre	Numerico	0-4
Fjob	Lavoro del padre	Nominale	Nota ²
Guardian	Adulto che sorveglia il ragazzo	Nominale	"father", "mother", "other"
famsize	Numero componenti famiglia	Binario	"GT3" (> 3), "LE3" (<= 3)
famrel	Qualità rapporti familiari	Numerico	1-5 (pessimi-ottimi)
reason	Motivo scelta scuola	Nominale	Nota ³
traveltime	Tempo necessario per arrivare a scuola	Numerico	1-4 ⁴
studytime	Ore studio settimanali	Numerico	1-4 ⁵
failures	Numero di bocciature	Numerico	n se $1 \leq n \leq 3$, altrimenti 4
schoolsup	Supporto extra scolastico	Binario	"Yes", "No"

¹0-nessuna, 1-educazione primaria (4 classi), 2-media (5-9 classi), 3-secondaria, 4-università

²"at-home", "health", "services", "teacher", "other"

³"course" (ottimi corsi), "home" (vicino casa), "reputation" (ottima reputazione), "other"

⁴1- < 15 min, 2-da 15 a 30 min, 3-da 30 min a 1 ora, 4-oltre 1 ora

⁵1- < 2h, 2-da 2 a 5h, 3-da 5 a 10h, 4-oltre 10h

famsup	Supporto familiare	Binario	"Yes", "No"
activities	Attività extracurricolari	Binario	"Yes", "No"
paidclass	Corsi addizionali	Binario	"Yes", "No"
internet	Accesso ad internet	Binario	"Yes", "No"
nursery	Asilo Nido	Binario	"Yes", "No"
higher	Desiderio di proseguire gli studi	Binario	"Yes", "No"
romantic	Ha un/una ragazzo/a?	Binario	"Yes", "No"
freetime	Tempo Libero	Numerico	1-5 (poco-molto)
goout	Uscite con gli amici	Numerico	1-5 (poche-molte)
Walc	Consumo alcohol nel weekend	Numerico	1-5 (poco-molto)
Dalc	Consumo alcohol in settimana	Numerico	1-5 (poco-molto)
health	Stato di salute	Numerico	1-5 (non buono-eccellente)
absences	Assenze a scuola	Numerico	0-93
G1	Voto scuola primaria	Numerico	0-20
G2	Voto scuola media	Numerico	0-20
G3	Voto scuola superiore	Numerico	0-20