

Cognome e Nome:**Matricola:****Esercizio 1 (punti 6 su 30)**

Disegnare un percettrone multilivello, con indicazione dei pesi e delle step function, che calcoli la funzione booleana OR.

Esercizio 2 (punti 6 su 30)

Data una tabella con 6 tuple. Si supponga che su un attributo X valgano le seguenti relazioni:

$$t_1[X] = t_2[X] \neq t_3[X] = t_4[X] \neq t_5[X] = t_6[X] \approx t_1[X]$$

$$t_1[Y] = t_2[Y] \neq t_3[Y] \approx t_4[Y] \neq t_5[Y] = t_6[Y] \approx t_1[Y]$$

dove il simbolo \approx indica che è possibile trovare una soglia ragionevole affinché due tuple siano simili, mentre ciò non vale per il simbolo \neq .

- a) Dire se esiste una FD e/o una RFD $X \rightarrow Y$. Qualora valga una RFD che rilassi sull'extent, indicare il relativo g3 error.
- b) Modificare le relazioni di cui sopra in modo che possa valere sia una FD che una RFD

Esercizio 3 (punti 6 su 30)

Utilizzando il paradigma Map Reduce, scrivere le funzioni Map e Reduce di un programma che, letto in input un file contenente cifre, emetta in output la frequenza con cui ciascun digit compreso tra 0 e 9 appare nel file di input.

Esercizio 4 (punti 8 su 30)

Dato il seguente training set relativo a sofferenze bancarie di debitori che hanno mancato il pagamento di alcune rate di un prestito bancario (Sofferenza = SI):

Soggetto	Stipendio	Titolo Studio	Stato Civile	Sofferenza
1	0-15K	Diploma	Celibe	NO
2	15-30K	Laurea	Coniugato	NO
3	15-30K	Diploma	Separato	SI
4	0-15K	< Diploma	Coniugato	SI
5	30-50K	Dottorato	Coniugato	NO
6	>50K	Laurea	Separato	NO
7	30-50K	Laurea	Separato	SI
8	15-30K	Laurea	Coniugato	SI
9	15-30K	Diploma	Coniugato	NO
10	15-30K	Diploma	Separato	SI

- a) Costruire un albero di decisione usando l'indice Gini ed indicando in ogni nodo il valore dell'indice e la distribuzione dei campioni;
- b) Come a) usando l'entropia;

Esercizio 5 (punti 4 su 30)

Trasformare l'attributo categorico *Stato Civile* dell'esercizio 4 in attributo numerico utilizzando la tecnica One Hot Encoding.