

FD/RFD

Esercizio 1

Data una tabella con 6 tuple, si supponga che sugli attributi X e Y valgano le seguenti relazioni:

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

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

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 .

- Dire se esiste una FD e/o una RFD $X \rightarrow Y$.
- Qualora valga una RFD che rilassi sull'extent, indicare il relativo g3error.

SOLUZIONE:

costruire un esempio pratico può aiutarvi

- Non esiste una FD $X \rightarrow Y$ ($t_4[X] = t_5[X]$ ma $t_4[Y] \approx t_5[Y]$)
- Considerate queste relazioni, vale la seguente AFD:

$$D_{\text{TRUE}} : X_{\text{EQ}} \xrightarrow{\psi(X,Y) \leq 1/6} Y_{\text{EQ}}$$

	X	Y
t ₁	1	10
t ₂	2	20
t ₃	2	20
t ₄	10	20
t ₅	10	21
t ₆	1	10

Minhashing

Esercizio 2

Shingle	S ₁	S ₂	S ₃	S ₄	S ₅
0	1	1	0	0	0
1	0	1	1	1	1
2	0	0	0	1	0
3	0	1	0	0	1
4	1	1	0	1	1
5	0	0	1	1	0
6	1	0	1	0	1
7	0	0	1	1	0

- a) Calcolare la similarità di Jaccard tra ogni coppia di colonne
- b) Calcolare le signature di ogni colonna usando le seguenti 6 funzioni hash:

$$h1(x) = (4x^2 + 3) \bmod 8; h2(x) = (2x^3 + 7) \bmod 8; h3(x) = (3x^4 + 5) \bmod 8;$$

$$h4(x) = (5x^2 - 2) \bmod 8; h5(x) = (7x - 3) \bmod 8; h6(x) = (6x^3 + 8) \bmod 8;$$

- c) Sulla base delle similarità di Jaccard calcolate al punto b), per ogni coppia di colonne calcolare la probabilità che la coppia venga selezionata per il confronto, ipotizzando che la matrice delle signature sia partizionata in 2 bande di 3 righe ciascuna.

SOLUZIONE:

a) Denominatore: #shingle che hanno almeno un 1 su S_x e S_y

Numeratore: #shingle che hanno 1 su entrambi S_x e S_y

$$\text{SIM}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 2/5$$

$$\text{SIM}(S_2, S_4) = 2/7$$

$$\text{SIM}(S_1, S_3) = 1/6$$

$$\text{SIM}(S_2, S_5) = 3/5$$

$$\text{SIM}(S_1, S_4) = 1/7$$

$$\text{SIM}(S_3, S_4) = 3/6$$

$$\text{SIM}(S_1, S_5) = 2/5$$

$$\text{SIM}(S_3, S_5) = 2/6$$

$$\text{SIM}(S_2, S_3) = 1/7$$

$$\text{SIM}(S_4, S_5) = 2/7$$

b)

STATO INIZIALE

	S1	S2	S3	S4	S5
H1	∞	∞	∞	∞	∞
H2	∞	∞	∞	∞	∞
H3	∞	∞	∞	∞	∞
H4	∞	∞	∞	∞	∞
H5	∞	∞	∞	∞	∞
H6	∞	∞	∞	∞	∞

PASSO 0

$$H1(0) = (4 \cdot 0^2 + 3) \bmod 8 = 3$$

$$H2(0) = 7$$

$$H3(0) = 5$$

$$H4(0) = 6$$

$$H5(0) = 5$$

$$H6(0) = 0$$

	S1	S2	S3	S4	S5
H1	3	3	∞	∞	∞
H2	7	7	∞	∞	∞
H3	5	5	∞	∞	∞
H4	6	6	∞	∞	∞
H5	5	5	∞	∞	∞
H6	0	0	∞	∞	∞

PASSO 1

$$H1(1) = 7$$

$$H2(1) = 1$$

$$H3(1) = 0$$

$$H4(1) = 3$$

$$H5(1) = 4$$

$$H6(1) = 6$$

	S1	S2	S3	S4	S5
H1	3	3	7	7	7
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	6	3	3	3	3
H5	5	4	4	4	4
H6	0	0	6	6	6

PASSO 2

$$H1(2) = 3$$

$$H2(2) = 7$$

$$H3(2) = 5$$

$$H4(2) = 2$$

$$H5(2) = 3$$

$$H6(2) = 0$$

	S1	S2	S3	S4	S5
H1	3	3	7	3	7
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	6	3	3	2	3
H5	5	4	4	3	4
H6	0	0	6	0	6

PASSO 3

$$H1(3) = 7$$

$$H2(3) = 5$$

$$H3(3) = 6$$

$$H4(3) = 3$$

$$H5(3) = 2$$

$$H6(3) = 2$$

	S1	S2	S3	S4	S5
H1	3	3	7	3	7
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	6	3	3	2	3
H5	5	2	4	3	2
H6	0	0	6	0	2

PASSO 4

$$H1(4) = 3$$

$$H2(4) = 7$$

$$H3(4) = 5$$

$$H4(4) = 6$$

$$H5(4) = 1$$

$$H6(4) = 0$$

	S1	S2	S3	S4	S5
H1	3	3	7	3	3
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	6	3	3	2	3
H5	1	1	4	1	1
H6	0	0	6	0	0

PASSO 5

$$H1(5) = 7$$

$$H2(5) = 1$$

$$H3(5) = 0$$

$$H4(5) = 3$$

$$H5(5) = 0$$

$$H6(5) = 6$$

	S1	S2	S3	S4	S5
H1	3	3	7	3	3
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	6	3	3	2	3
H5	1	1	0	0	1
H6	0	0	6	0	0

PASSO 6

$$H1(6) = 3$$

$$H2(6) = 7$$

$$H3(6) = 5$$

$$H4(6) = 2$$

$$H5(6) = 7$$

$$H6(6) = 0$$

	S1	S2	S3	S4	S5
H1	3	3	3	3	3
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	2	3	2	2	2
H5	1	1	0	0	1
H6	0	0	0	0	0

PASSO 7

$$H1(7) = 7$$

$$H2(7) = 5$$

$$H3(7) = 0$$

$$H4(7) = 3$$

$$H5(7) = 6$$

$$H6(7) = 2$$

	S1	S2	S3	S4	S5
H1	3	3	3	3	3
H2	7	1	1	1	1
H3	5	0	0	0	0
H4	2	3	2	2	2
H5	1	1	0	0	1
H6	0	0	0	0	0

Denominatore: # righe della matrice

Numeratore: # righe che hanno valori uguali su S_x e S_y

$$SIM(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 3/6$$

$$SIM(S_2, S_4) = 4/6$$

$$SIM(S_1, S_3) = 3/6$$

$$SIM(S_2, S_5) = 5/6$$

$$SIM(S_1, S_4) = 3/6$$

$$SIM(S_3, S_4) = 6/6$$

$$SIM(S_1, S_5) = 4/6$$

$$SIM(S_3, S_5) = 5/6$$

$$SIM(S_2, S_3) = 4/6$$

$$SIM(S_4, S_5) = 5/6$$

c)

Formula: $1 - (1 - SIM^r)^b$

$$r = 3 \quad b = 2$$

$$Prob(S1, S2) = 1 - (1 - 0.5^3)^2 = 0.23$$

$$Prob(S1, S3) = 0.23$$

$$Prob(S2, S4) = 0.51$$

$$Prob(S1, S4) = 0.23$$

$$Prob(S2, S5) = 0.82$$

$$Prob(S1, S5) = 0.51$$

$$Prob(S3, S4) = 1$$

$$Prob(S2, S3) = 0.51$$

$$Prob(S3, S5) = 0.82$$

$$Prob(S4, S5) = 0.82$$

Map Reduce

Esercizio 3

Scrivere le funzioni Map e Reduce di un programma distribuito che legga in input un database di calciatori e calcoli, per ogni squadra, il numero di calciatori che hanno militato nella stessa.

SOLUZIONE:

MAP (Key, Value)

```
//Key = chunk del database
//Value = contenuto del chunk
For each (Calciatore c in Value)
    s = c.getSquadra()
    Emit (s,1)
```

REDUCE (Key, Values)

```
//Key = una squadra
//Values = iteratore sui conteggi
result = 0
For each (count c in Values)
    result = result + c
    Emit (Key,result)
```

Data Integration

Esercizio 4

Dati i seguenti due schemi di database, di cui DB1 gestito dalla Federazione Italiana Gioco Calcio (FIGC) e DB2 gestito dalla Lega Calcio di serie A:

// DB1 FIGC

SQUADRE2021(Nome, Calciatore, Data-Inizio, Data-Fine, Ingaggio)

CALCIATORE(N.Tessera, Cognome, Nome, Nazionalità, Data-Nascita, Statura, Ruolo)

// DB2 Lega Calcio

CALCIATORE(Num-Tessera, Codice Fiscale, Cognome, Nome, Email, Goal in carriera)

SQUADRE2022(Nome, Calciatore, Data-Inizio, Data-Fine, Ingaggio)

Integrare i due schemi tramite una vista ed utilizzando l'approccio GAV.

SOLUZIONE:

```
Create view Calciatori as (Select DB1FIGC.Calciatore.Nome,
                                .Cognome,
                                .N.tessera,
                                .Nazionalità,
                                .Data nascita,
                                .Statura,
                                .Ruolo,
                                DB2LegaCalcio.Calciatore.Codice Fiscale,
                                .email,
                                .gol in carriera
                                From DB1FIGC.Calciatore
                                JOIN
                                DB2LegaCalcio.Calciatore
                                ON
                                DB1FIGC.Calciatore.Ntessera =
                                DB2LegaCalcio.Calciatore.Num_Tessera
                                )
```

```
Create view Squadre as (Select * From DB1FIGC.Squadre2021
                        UNION
                        DB2LegaCalcio.Squadre2022
                        )
```

Data Integration

Esercizio 5

Dati i seguenti due schemi di database, di cui DB1 gestito dal nostro Dipartimento (Informatica) e DB2 gestito dalla segreteria studenti:

// DB1 Dipartimento

ESAMI2021(Codice Corso, Studente, Data, Voto, Lode)

STUDENTE(Matricola, Cognome, Nome, Indirizzo, Cap, Città, Provincia)

// DB2 Segreteria

STUDENTE(Matricola, Codice Fiscale, Cognome, Nome, Telefono, CFU Superati)

ESAMI2022(Codice Corso, Studente, Data, Voto, Lode)

- a) Integrare i due schemi tramite una vista ed utilizzando l'approccio GAV.
- b) Riferendosi alla soluzione di a), integrare lo schema globale tramite una vista ed utilizzando un approccio LAV (esamiLode/esamiSenzaLode) (studentiAV/studentiSA).

SOLUZIONE:

a)

```
Create view Studenti as (Select DB1.Studente.Matricola,  
                                .Cognome,  
                                .Nome,  
                                .Indirizzo,  
                                .CAP,  
                                .Città,  
                                .Provincia,  
                                DB2.Studente.Codice Fiscale,  
                                .telefono,  
                                .CFU  
From DB1.Studente  
      JOIN  
      DB2.Studente  
      ON  
      DB1.studente.matricola =  
      DB2.studente.matricola  
)
```

```
Create view Esami as (Select * From DB1.Esame2021  
                        UNION  
                        DB2.Esami2022  
)
```

b)

```
Create view EsamiConLode as (Select * From DB.Esami  
                             Where lode ="SI"  
                             )
```

```
Create view EsamiSenzaLode as (Select * From DB.Esami  
                               Where lode ="NO"  
                               )
```

```
Create view StudentiAV as (Select * From DB.Studenti  
                           Where provincia ="AV"  
                           )
```

```
Create view StudentiSA as (Select * From DB.Studenti  
                           Where provincia ="SA"  
                           )
```

*Assumendo che nel nostro DB abbiamo solo studenti di AV e di SA.

Correlazione di Pearson

Esercizio 6

X₁	X₂	X₃	X₄	Y
2	3	-1	-2	0
1	6	4	5	1
-5	2	-1	-4	0
-7	3	-5	4	1
-10	8	-9	6	1
-18	17	-12	2	0
1	5	-10	7	1
6	-5	-8	3	0

Usando il coefficiente di Pearson, determinare l'attributo predittivo che sia maggiormente correlato all'attributo dipendente Y. La formula è la seguente

$$PC = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Dove n nel nostro caso è 8

\bar{x} è la media di tutti i valori di una variabile X_i

s è la deviazione standard: $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

SOLUZIONE:

calcolo delle medie:

$$\text{avg}(x_1) = -30/8 = -3.75 \quad \text{avg}(x_2) = 39/8 = 4.875$$

$$\text{avg}(x_3) = -42/8 = -5.25 \quad \text{avg}(x_4) = 21/8 = 2.625$$

$$\text{avg}(y) = 4/8 = 0.5$$

Calcolo delle deviazioni standard:

$$\text{Formula: } s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \text{avg}(x))^2}$$

$$S_{x1} = 7.81 \quad S_{x2} = 6.22 \quad S_{x3} = 5.50 \quad S_{x4} = 3.85 \quad S_y = 0.53$$

calcolo dei coefficienti di Pearson:

$$\text{Formula: } PC = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \text{avg}(x)}{s_x} \right) \left(\frac{y_i - \text{avg}(y)}{s_y} \right)$$

$$\begin{aligned} PC(x_1) = \frac{1}{7} & \left[\left(\frac{2+3.75}{7.81} \right) \left(\frac{0-0.5}{0.53} \right) + \left(\frac{1+3.75}{7.81} \right) \left(\frac{1-0.5}{0.53} \right) + \right. \\ & \left(\frac{-5+3.75}{7.81} \right) \left(\frac{0-0.5}{0.53} \right) + \left(\frac{-7+3.75}{7.81} \right) \left(\frac{1-0.5}{0.53} \right) + \\ & \left(\frac{-10+3.75}{7.81} \right) \left(\frac{1-0.5}{0.53} \right) + \left(\frac{-18+3.75}{7.81} \right) \left(\frac{0-0.5}{0.53} \right) + \\ & \left. \left(\frac{1+3.75}{7.81} \right) \left(\frac{1-0.5}{0.53} \right) + \left(\frac{6+3.75}{7.81} \right) \left(\frac{0-0.5}{0.53} \right) \right] \end{aligned}$$

$$= \frac{1}{7} [(0.74)(-0.94) + (0.61)(0.94) + (-0.16)(-0.94) + (-0.42)(0.94) + (-0.8)(0.94) + (-1.82)(-0.94) + (0.61)(0.94) + (1.25)(-0.94)]$$

$$= \frac{1}{7} * 0.02 \approx 0$$

$$PC(x_2) = 0.107$$

$$PC(x_3) = 0.05$$

$$PC(x_4) = 0.8$$

One Hot Encoding

Esercizio 7

Soggetto	Età	Anni come Cliente	Spesa Anno Precedente	Cambiato Fornitore
1	30-50	1-5	601-800 €	NO
2	> 50	5-10	801-1200 €	SI
3	30-50	> 10	301-600 €	NO
4	24-29	1-5	100-300 €	SI
5	30-50	5-10	301-600 €	NO
6	24-29	> 10	100-300 €	SI
7	24-29	1-5	100-300 €	NO
8	30-50	5-10	801-1200 €	NO
9	> 50	> 10	301-600 €	SI
10	24-29	1-5	100-300 €	SI

Dato questo dataset, applicare la tecnica dell'one hot encoding sull'attributo categorico Età.

SOLUZIONE:

Soggetto	24 - 29	30 – 50	>50	Anni c. c.	Spesa a. p.	Cambio F.
1	0	1	0	1 – 5	601 – 800	No
2	0	0	1	5 – 10	801 – 1200	Si
3	0	1	0	>10	301 – 600	No
4	1	0	0	1 – 5	100 - 300	Si
5	0	1	0	5 – 10	301 – 600	No
6	1	0	0	>10	100 - 300	Si
7	1	0	0	1 – 5	100 - 300	No
8	0	1	0	5 – 10	801 – 1200	No
9	0	0	1	>10	301 – 600	Si
10	1	0	0	1 – 5	100 - 300	Si

Normalizzazione

Esercizio 8

Supponiamo di avere un dataset che ha una feature contenente le altezze di una serie di persone in centimetri:

Altezze: 160, 170, 155, 180, 165, 175, 150

Normalizzare queste altezze utilizzando la tecnica di Min-Max Scaling per assicurare che siano comprese in un intervallo tra 0 e 1.

SOLUZIONE:

$$\text{Formula Min-Max} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

$$\text{Max} = 180$$

$$\text{Min} = 150$$

$$160 \rightarrow \frac{160 - 150}{180 - 150} = 0.33$$

$$170 \rightarrow \frac{170 - 150}{180 - 150} = 0.67$$

$$155 \rightarrow \frac{155 - 150}{180 - 150} = 0.17$$

$$180 \rightarrow \frac{180 - 150}{180 - 150} = 1$$

$$165 \rightarrow \frac{165 - 150}{180 - 150} = 0.5$$

$$175 \rightarrow \frac{175 - 150}{180 - 150} = 0.84$$

$$150 \rightarrow \frac{150 - 150}{180 - 150} = 0$$

Similarity Features

Esercizio 9

Si consideri il seguente dataset monodimensionale, che comprende due classi (verde e rosso):

-4	-2	-1	2	3	3.5	4	8.5
----	----	----	---	---	-----	---	-----

Calcolare, attraverso il metodo RBF (Radial Basis Function), il valore di due nuove feature per i punti $x_1 = 3$ $x_2 = 3.5$, considerando come landmark i punti $l_1 = 2$ $l_2 = 4$ e come parametro $\gamma = 0.4$.

SOLUZIONE:

Formula $= \phi_\gamma(x, l) = \exp(-\gamma \|x - l\|^2)$

$$l_1 = 2 \quad l_2 = 4 \quad x_1 = 3 \quad x_2 = 3.5 \quad \gamma = 0.4$$

$$\phi_\gamma(x_1, l_1) = \exp(-\gamma \|x_1 - l_1\|^2)$$

$$\phi_{0.4}(3, 2) = \exp(-0.4 \|3 - 2\|^2)$$

$$= \exp(-0.4 * 1^2)$$

$$= \exp(-0.4) \approx 0.67$$

$$\phi_\gamma(x_1, l_2) = \exp(-\gamma \|x_1 - l_2\|^2)$$

$$\phi_{0.4}(3, 4) = \exp(-0.4 \|3 - 4\|^2)$$

$$= \exp(-0.4 * 1^2)$$

$$= \exp(-0.4) \approx 0.67$$

$$\phi_\gamma(x_2, l_1) = \exp(-\gamma \|x_2 - l_1\|^2)$$

$$\phi_{0.4}(3.5, 2) = \exp(-0.4 \|3.5 - 2\|^2)$$

$$= \exp(-0.4 * 1.5^2)$$

$$= \exp(-0.4 * 2.25)$$

$$= \exp(-0.4 * 2.25) \approx 0.40$$

$$\phi_\gamma(x_2, l_2) = \exp(-\gamma \|x_2 - l_2\|^2)$$

$$\phi_{0.4}(3.5, 4) = \exp(-0.4 \|3.5 - 4\|^2)$$

$$= \exp(-0.4 * 1^2)$$

$$= \exp(-0.4 * 2.25)$$

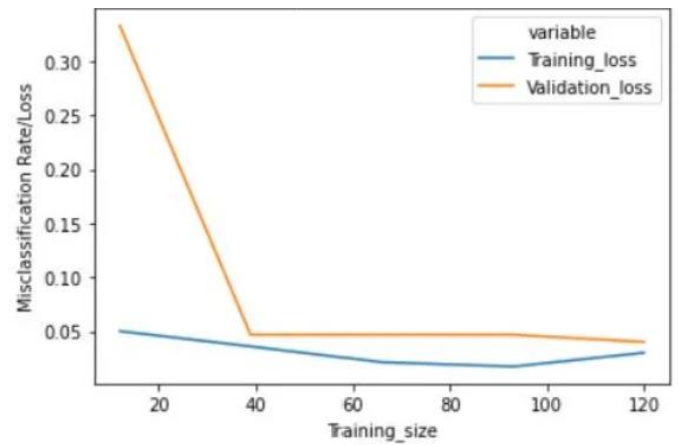
$$= \exp(-0.4 * 2.25) \approx 0.90$$

Learning Curve

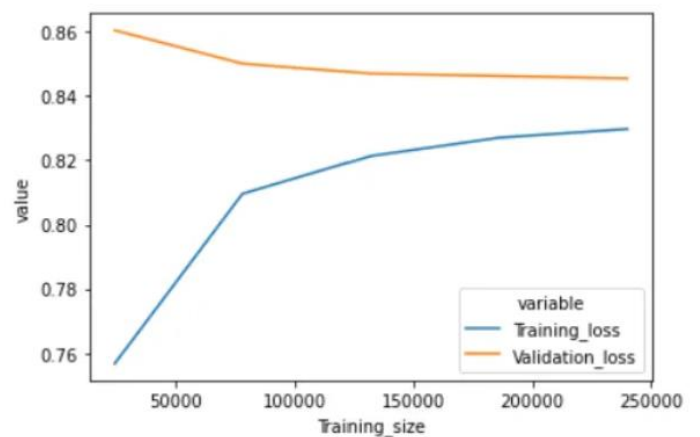
Esercizio 10

Si considerino le seguenti curve di apprendimento relative a tre modelli. Determinare:

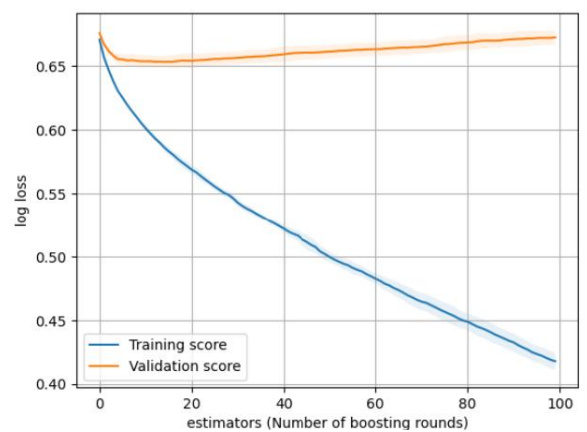
- Quale modello va in overfitting
- Quale modello va in underfitting
- Quale modello mostra una buona curva di apprendimento



A



B



C

SOLUZIONE:

- Overfitting è modello C
- Underfitting è modello B
- Il modello con una buona curva di apprendimento è A

IMPORTANTE: motivate la risposta!

Equazione Normale e funzione sigmoide

Esercizio 11

Si consideri il seguente dataset, dove gli attributi X_1 , X_2 e X_3 rappresentano le feature di ogni istanza, mentre l'attributo Y contiene il valore della classe assunta (classificazione binaria, classi 0 o 1).

X_1	X_2	X_3	Y
1	4	1	1
3	3	1	1
2	1	3	0
4	0	4	0

- a) Utilizzare la Normal Equation per calcolare i pesi relativi ad ogni attributo X_i ed il termine di Bias.
- b) Supponiamo di avere le seguenti due istanze:

X_1	X_2	X_3
2	-3	5
2	3	5
0	4	2

Utilizzando i pesi definiti al punto a), calcolare le predizioni delle seguenti istanze. Inoltre, per ogni predizione ottenuta, applicare la funzione Sigmoide per identificare la classe di ogni istanza.

SOLUZIONE:

a)

formula per calcolare pesi e bias tramite eq. Normale:

$$\theta = (X^T X)^{-1} X^T Y$$

dove X è il dataset (senza Y) con una colonna aggiuntiva per il bias (tutti i valori settati ad 1)

$$X = \begin{pmatrix} 1 & 4 & 1 & 1 \\ 3 & 3 & 3 & 1 \\ 2 & 1 & 3 & 1 \\ 4 & 0 & 4 & 1 \end{pmatrix} \quad Y = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 3 & 2 & 4 \\ 4 & 3 & 1 & 0 \\ 1 & 1 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 4 & 1 & 1 \\ 3 & 3 & 1 & 1 \\ 2 & 1 & 3 & 1 \\ 4 & 0 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 15 & 26 & 10 \\ 15 & 26 & 10 & 8 \\ 26 & 10 & 27 & 9 \\ 10 & 8 & 9 & 4 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & 3 & 2 & 4 \\ 4 & 3 & 1 & 0 \\ 1 & 1 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 2 \\ 2 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 30 & 15 & 26 & 10 \\ 15 & 26 & 10 & 8 \\ 26 & 10 & 27 & 9 \\ 10 & 8 & 9 & 4 \end{pmatrix}^{(-1)} = \begin{pmatrix} \frac{14}{25} & \frac{23}{25} & \frac{4}{5} & \frac{-126}{25} \\ \frac{23}{25} & \frac{86}{25} & \frac{18}{5} & \frac{-432}{25} \\ \frac{4}{5} & \frac{18}{5} & 4 & \frac{-91}{5} \\ \frac{-126}{25} & \frac{-432}{25} & \frac{-91}{5} & \frac{2209}{25} \end{pmatrix}$$

$$(X^T X)^{-1} X^T Y = \begin{pmatrix} \frac{14}{25} & \frac{23}{25} & \frac{4}{5} & \frac{-126}{25} \\ \frac{23}{25} & \frac{86}{25} & \frac{18}{5} & \frac{-432}{25} \\ \frac{4}{5} & \frac{18}{5} & 4 & \frac{-91}{5} \\ \frac{-126}{25} & \frac{-432}{25} & \frac{-91}{5} & \frac{2209}{25} \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 7 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \\ \frac{2}{5} \\ 0 \\ \frac{-4}{5} \end{pmatrix}$$

Pesi: $w_1 = 1/5$ $w_2 = 2/5$ $w_3 = 0$ $b = -4/5$

b)

Formula predizione: $x_1 w_1 + x_2 w_2 + x_3 w_3 + b$

$$t_1 = (2)(1/5) + (-3)(2/5) + (5)(0) - 4/5 = -1.6$$

$$t_2 = (2)(1/5) + (3)(2/5) + (5)(0) - 4/5 = 0.8$$

$$t_3 = (0)(1/5) + (4)(2/5) + (2)(0) - 4/5 = 0.8$$

Formula Sigmoidale: $\sigma(t) = \frac{1}{1 + \exp(-t)}$

$$\sigma(-1.6) = \frac{1}{1 + \exp(-(-1.6))} = 0.17 \quad \text{classe 0}$$

$$\sigma(0.8) = \frac{1}{1 + \exp(-0.8)} = 0.69 \quad \text{classe 1}$$

$$\sigma(0.8) = \frac{1}{1 + \exp(-0.8)} = 0.69 \quad \text{classe 1}$$

Classificazione

Esercizio 12

	True positive label	True negative label
Predicted Positive	15	45
Predicted Negative	97	77

Calcolare Precision, Recall, F1-score, Accuracy e True Negative Rate.

Soluzione:

$$\text{Precision} = \frac{TP}{TP+FP} \Rightarrow \frac{15}{15+45} = 0.25$$

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP+FN} \Rightarrow \frac{15}{15+97} = 0.1339$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \Rightarrow 2 * \frac{0.25 * 0.1339}{0.25 + 0.1339} = 0.17744$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \Rightarrow \frac{15+77}{15+45+77+97} = 0.3931$$

$$\text{True Negative Rate} = \text{Specificity} = \frac{TN}{TN+FP} \Rightarrow \frac{77}{77+45} = 0.6311$$

FD/RFD

Esercizio 1

Item	X ₁	X ₂	X ₃	X ₄	Y
1	2	3	-1	-2	0
2	1	6	4	5	1
3	-5	2	-1	-4	0
4	-7	3	-5	4	1
5	-10	8	-9	6	1
6	-18	17	-12	2	0
7	1	5	-10	7	1
8	6	-5	-8	3	0

Per ognuna delle 4 dipendenze candidate del tipo $X_i \rightarrow Y$, dire se vale come FD o, in caso contrario, dire se vale come RFD, nel qual caso specificare le soglie numeriche di similarità e/o le soglie per il g3 error, a seconda che si tratti di RFD che rilassa sul confronto, sull'extent, o su entrambi.

SOLUZIONE:

$D_{\text{TRUE}} : X_1 \longrightarrow Y$ vale come FD

$D_{\text{TRUE}} : X_2 \longrightarrow Y$ non vale come FD ma può valere come RFD:

$$D_{\text{TRUE}} : X_2 \text{ EQ } \xrightarrow{\psi(X,Y) \leq 0,125} Y_{\text{EQ}}$$

$D_{\text{TRUE}} : X_3 \longrightarrow Y$ vale come FD

$D_{\text{TRUE}} : X_4 \longrightarrow Y$ vale come FD

Esercizio 2

Tempo	Temperatura	Umidità	Vento	Gioca
Soleggiato	Calda	Alta	NO	<i>NO</i>
Soleggiato	Calda	Alta	SI	<i>NO</i>
Nuvoloso	Calda	Alta	NO	<i>SI</i>
Piovoso	Mite	Alta	NO	<i>SI</i>
Piovoso	Fredda	Normale	NO	<i>SI</i>
Piovoso	Fredda	Normale	SI	<i>NO</i>
Nuvoloso	Fredda	Normale	SI	<i>SI</i>
Soleggiato	Mite	Alta	NO	<i>NO</i>
Soleggiato	Fredda	Normale	NO	<i>SI</i>
Piovoso	Mite	Normale	NO	<i>SI</i>
Soleggiato	Mite	Normale	SI	<i>SI</i>
Nuvoloso	Mite	Alta	SI	<i>SI</i>
Nuvoloso	Calda	Normale	NO	<i>SI</i>
Piovoso	Mite	Alta	SI	<i>NO</i>

Facendo riferimento al dataset dell'esercizio 2

- (punti 2) Trovare eventuali FD con RHS singolo;
- (punti 4) Trovare eventuali RFD con RHS singolo che rilassano sull'extent e con un g3-error inferiore al 15%.

SOLUZIONE:

a)

$D_{TRUE} : \text{Tempo}, \text{Temperatura}, \text{Umidità}, \text{Vento} \longrightarrow \text{Gioca}$

$D_{TRUE} : \text{Tempo}, \text{Temperatura}, \text{Vento} \longrightarrow \text{Gioca}$

b)

$D_{TRUE} : \text{Umidità}_{EQ}, \text{Tempo}_{EQ}, \text{Temperatura}_{EQ} \xrightarrow{\psi(x,y) \leq 2/14} \text{Gioca}_{EQ}$

Esercizio 3

Data la seguente tabella:

Tuple number	Height	Weight	Shoe size
1	175	70	40
2	175	75	39
3	175	69	40
4	176	71	40
5	178	81	41
6	169	73	37
7	170	62	39

In relazione agli attributi *Height*, *Weight* e *Shoe size*:

- Elencare eventuali FD con lato destro a singolo attributo
- Elencare eventuali RFD con RHS singolo che rilassano solo sul confronto, solo sull'extent o su entrambi, mostrando le relative soglie di similarità o della misura di coverage g3 error

SOLUZIONE:

a)

$D_{TRUE} : \text{Weight} \longrightarrow \text{Shoe Size}$

$D_{TRUE} : \text{Weight} \longrightarrow \text{Height}$

b)

$D_{TRUE} : \text{Height}_{EQ} \xrightarrow{\Psi(X,Y) \leq 2/7} \text{Weight}_{EQ}$

$D_{TRUE} : \text{Height}_{EQ} \xrightarrow{\Psi(X,Y) \leq 1/7} \text{Shoe Size}_{EQ}$

$$\begin{aligned}
 D_{TRUE} : \text{Height}_{EQ} &\xrightarrow{\Psi_{\text{ext}}(0)} \text{Shoe Size}_{EQ} \quad \max_{\lambda \in \Pi_X} \Delta\Phi(\lambda[Y]) \leq 1 \\
 D_{TRUE} : \text{Shoe Size}_{EQ} &\xrightarrow{\Psi_{\text{ext}}(0)} \text{Weight}_{EQ} \quad \max_{\lambda \in \Pi_X} \Delta\Phi(\lambda[Y]) \leq 1 \\
 D_{TRUE} : \text{Height}_{EQ} &\xrightarrow{\Psi(X,Y) = 1/7} \text{Weight}_{EQ} \quad \max_{\lambda \in \Pi_X} \Delta\Phi(\lambda[Y]) \leq 1
 \end{aligned}$$

MINHASHING

Esercizio 4

Shingle	S ₁	S ₂	S ₃	S ₄
0	1	0	0	0
1	0	0	1	1
2	0	1	0	0
3	0	1	1	0
4	1	0	1	1
5	1	0	0	1

- a) Calcolare la similarità di Jaccard tra ogni coppia di colonne;
- b) Calcolare le signature di ogni colonna usando le seguenti 4 funzioni hash:

$$h1(x) = (4x^2 + 1) \bmod 6$$

$$h2(x) = (x^3 + 2) \bmod 6$$

$$h3(x) = (x^4 + 3) \bmod 6$$

$$h4(x) = (2x + 4) \bmod 6$$

Mostrare l'evoluzione della matrice delle signature di minhash, simulando l'esecuzione dell'algoritmo per il loro calcolo. Inoltre, calcolare le similarità di Jaccard tra tutte le coppie di signature di minhash.

- c) Sulla base delle similarità di Jaccard calcolate al punto b), per ogni coppia di colonne calcolare la probabilità che la coppia venga selezionata per il confronto, ipotizzando che la matrice delle signature sia partizionata in 2 bande di 3 righe ciascuna.

SOLUZIONE:

a) Denominatore: #shingle che hanno almeno un 1 su S_x e S_y

Numeratore: #shingle che hanno 1 su entrambi S_x e S_y

$$\text{SIM}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 0/5$$

$$\text{SIM}(S_1, S_3) = 1/5$$

$$\text{SIM}(S_1, S_4) = 1/2$$

$$\text{SIM}(S_2, S_3) = 1/4$$

$$\text{SIM}(S_2, S_4) = 0/5$$

$$\text{SIM}(S_3, S_4) = 1/2$$

b)

STATO INIZIALE

	S1	S2	S3	S4
H1	∞	∞	∞	∞
H2	∞	∞	∞	∞
H3	∞	∞	∞	∞
H4	∞	∞	∞	∞

PASSO 0

$$H1(0) = (4 \cdot 0^2 + 1) \bmod 6 = 1$$

$$H2(0) = 2$$

$$H3(0) = 3$$

$$H4(0) = 4$$

	S1	S2	S3	S4
H1	1	∞	∞	∞
H2	2	∞	∞	∞
H3	3	∞	∞	∞
H4	4	∞	∞	∞

PASSO 1

$$H1(1) = 5$$

$$H2(1) = 3$$

$$H3(1) = 4$$

$$H4(1) = 0$$

	S1	S2	S3	S4
H1	1	∞	5	5
H2	2	∞	3	3
H3	3	∞	4	4
H4	4	∞	0	0

PASSO 2

$$H1(2) = 5$$

$$H2(2) = 4$$

$$H3(2) = 1$$

$$H4(2) = 2$$

	S1	S2	S3	S4
H1	1	5	5	5
H2	2	4	3	3
H3	3	1	4	4
H4	4	2	0	0

PASSO 3

$$H1(3) = 1$$

$$H2(3) = 5$$

$$H3(3) = 0$$

$$H4(3) = 4$$

	S1	S2	S3	S4
H1	1	1	1	5
H2	2	4	3	3
H3	3	0	0	4
H4	4	2	0	0

PASSO 4

$$H1(4) = 5$$

$$H2(4) = 0$$

$$H3(4) = 1$$

$$H4(4) = 0$$

	S1	S2	S3	S4
H1	1	1	1	5
H2	0	4	0	0
H3	1	0	0	1
H4	0	2	0	0

PASSO 5

$$H1(5) = 5$$

$$H2(5) = 1$$

$$H3(5) = 4$$

$$H4(5) = 2$$

	S1	S2	S3	S4
H1	1	1	1	5
H2	0	4	0	0
H3	1	0	0	1
H4	0	2	0	0

Denominatore: # righe della matrice

Numeratore: # righe che hanno valori uguali su S_x e S_y

$$\text{SIM}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 1/4$$

$$\text{SIM}(S_1, S_3) = 3/4$$

$$\text{SIM}(S_1, S_4) = 3/4$$

$$\text{SIM}(S_2, S_3) = 1/2$$

$$\text{SIM}(S_2, S_4) = 0/4$$

$$\text{SIM}(S_3, S_4) = 1/2$$

c)

Formula: $1 - (1 - \text{SIM}^r)^b$

$$r = 3$$

$$b = 2$$

$$\text{Prob}(S1, S2) = 1 - (1 - 0.25^3)^2 = 0.03$$

$$\text{Prob}(S1, S3) = 0.66$$

$$\text{Prob}(S1, S4) = 0.66$$

$$\text{Prob}(S2, S3) = 0.23$$

$$\text{Prob}(S2, S4) = 0$$

$$\text{Prob}(S3, S4) = 0.23$$

Esercizio 5

Shingle	S ₁	S ₂	S ₃	S ₄
0	1	1	0	1
1	0	0	1	1
2	1	1	0	0
3	0	0	1	0
4	0	1	1	0
5	1	0	0	0

- a) Calcolare la similarità di Jaccard tra ogni coppia di colonne;
- b) Calcolare la signature di ogni colonna usando le seguenti 4 funzioni hash:

$$h_1(x) = (x^4 + 1) \bmod 6$$

$$h_2(x) = (x^2 + 3) \bmod 6$$

$$h_3(x) = (x^3 + 7) \bmod 6$$

$$h_4(x) = (3x^2 + 5) \bmod 6$$

Mostrare l'evoluzione della matrice delle signature di minhash simulando l'esecuzione dell'algoritmo per il loro calcolo. Inoltre, calcolare le similarità di Jaccard tra tutte le coppie di signature di minhash.

C) (2 punti) Sulla base delle similarità di Jaccard calcolate al punto b), calcolare per ogni coppia di colonne la probabilità che la coppia venga selezionata per il confronto, ipotizzando che la matrice delle signature sia partizionata in 2 bande di 2 righe ciascuna.

SOLUZIONE:

a) Denominatore: #shingle che hanno almeno un 1 su S_x e S_y

Numeratore: #shingle che hanno 1 su entrambi S_x e S_y

$$\text{SIM}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 1/2$$

$$\text{SIM}(S_1, S_3) = 0$$

$$\text{SIM}(S_1, S_4) = 1/4$$

$$\text{SIM}(S_2, S_3) = 1/5$$

$$\text{SIM}(S_2, S_4) = 1/4$$

$$\text{SIM}(S_3, S_4) = 1/4$$

b)

STATO INIZIALE

	S1	S2	S3	S4
H1	∞	∞	∞	∞
H2	∞	∞	∞	∞
H3	∞	∞	∞	∞
H4	∞	∞	∞	∞

PASSO 0

$$H1(0) = (0^4 + 1) \bmod 6 = 1$$

$$H2(0) = 3$$

$$H3(0) = 1$$

$$H4(0) = 5$$

	S1	S2	S3	S4
H1	1	1	∞	1
H2	3	3	∞	3
H3	1	1	∞	1
H4	5	5	∞	5

PASSO 1

$$H1(1) = 2$$

$$H2(1) = 4$$

$$H3(1) = 2$$

$$H4(1) = 2$$

	S1	S2	S3	S4
H1	1	1	2	1
H2	3	3	4	3
H3	1	1	2	1
H4	5	5	2	2

PASSO 2

$$H1(2) = 5$$

$$H2(2) = 1$$

$$H3(2) = 3$$

$$H4(2) = 5$$

	S1	S2	S3	S4
H1	1	1	2	1
H2	1	1	4	3
H3	1	1	2	1
H4	5	5	2	2

PASSO 3

$$H1(3) = 4$$

$$H2(3) = 0$$

$$H3(3) = 4$$

$$H4(3) = 2$$

	S1	S2	S3	S4
H1	1	1	2	1
H2	1	1	0	3
H3	1	1	2	1
H4	5	5	2	2

PASSO 4

$$H1(4) = 5$$

$$H2(4) = 1$$

$$H3(4) = 5$$

$$H4(4) = 5$$

	S1	S2	S3	S4
H1	1	1	2	1
H2	1	1	0	3
H3	1	1	2	1
H4	5	5	2	2

PASSO 5

$$H1(5) = 2$$

$$H2(5) = 4$$

$$H3(5) = 0$$

$$H4(5) = 2$$

	S1	S2	S3	S4
H1	1	1	2	1
H2	1	1	0	3
H3	0	1	2	1
H4	2	5	2	2

Denominatore: # righe della matrice

Numeratore: # righe che hanno valori uguali su S_x e S_y

$$SIM(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 1/2$$

$$SIM(S_1, S_3) = 1/4$$

$$SIM(S_1, S_4) = 1/2$$

$$SIM(S_2, S_3) = 0$$

$$SIM(S_2, S_4) = 1/2$$

$$SIM(S_3, S_4) = 1/4$$

c)

Formula: $1 - (1 - SIM^r)^b$

$$r = 2$$

$$b = 2$$

$$\text{Prob}(S1, S2) = 1 - (1 - 0.5^2)^2 = 0.43$$

$$\text{Prob}(S1, S3) = 0.12$$

$$\text{Prob}(S1, S4) = 0.43$$

$$\text{Prob}(S2, S3) = 0$$

$$\text{Prob}(S2, S4) = 0.43$$

$$\text{Prob}(S3, S4) = 0.12$$

MAP REDUCE

Esercizio 6

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.

SOLUZIONE:

MAP (Key, Value)

//Key = nome documento

//Value = testo del documento

For each (Digit d in Value)

Emit (d,1)

REDUCE (Key, Values)

//Key = un digit

//Values = iteratore sui conteggi

result = 0

For each (count c in Values)

result = result + c

Emit (Key,result)

Esercizio 7

Utilizzando il paradigma Map Reduce, scrivere le funzioni Map e Reduce di un programma che, letto in input un file contenente parole, emetta in output una lista di coppie in cui il primo elemento è un numero intero $k > 0$, mentre il secondo rappresenta il numero di parole del documento contenenti k caratteri.

SOLUZIONE:

MAP (Key, Value)

```
//Key = nome documento
//Value = testo del documento
For each (Word w in Value)
    l = length(w)
    Emit (l,1)
```

REDUCE (Key, Values)

```
//Key = lunghezza
//Values = iteratore sui conteggi
result = 0
For each (count c in Values)
    result = result + c
    Emit (Key,result)
```

DATA INTEGRATION

Esercizio 8

Dati i seguenti due schemi di database, di cui DB1 gestito dalla Federazione Italiana Gioco Calcio (FIGC) e DB2 gestito dalla Lega Calcio di serie A:

// DB1 FIGC

SQUADRE2021(Nome, Calciatore, Data-Inizio, Data-Fine, Ingaggio)

CALCIATORE(N.Tessera, Cognome, Nome, Nazionalità, Data-Nascita, Statura, Ruolo)

// DB2 Lega Calcio

CALCIATORE(Num-Tessera, Codice Fiscale, Cognome, Nome, Email, Goal in carriera)

SQUADRE2022(Nome, Calciatore, Data-Inizio, Data-Fine, Ingaggio)

Integrare i due schemi tramite una vista ed utilizzando l'approccio GAV.

SOLUZIONE:

```
Create view Calciatori as (Select DB1FIGC.Calciatore.Nome,
                                .Cognome,
                                .N.tessera,
                                .Nazionalità,
                                .Data nascita,
                                .Statura,
                                .Ruolo,
                                DB2LegaCalcio.Calciatore.Codice Fiscale,
                                .email,
                                .gol in carriera
                                From DB1FIGC.Calciatore
                                JOIN
                                DB2LegaCalcio.Calciatore
                                ON
                                DB1FIGC.Calciatore.Ntessera =
                                DB2LegaCalcio.Calciatore.Num_Tessera
                                )
```

```
Create view Squadre as (Select * From DB1FIGC.Squadre2021
                        UNION
                        DB2LegaCalcio.Squadre2022
                        )
```


Esercizio 9

Dati i seguenti due schemi di database, di cui DB1 gestito dal nostro Dipartimento (Informatica) e DB2 gestito dalla segreteria studenti:

// DB1 Dipartimento

ESAMI2021(Codice Corso, Studente, Data, Voto, Lode)

STUDENTE(Matricola, Cognome, Nome, Indirizzo, Cap, Città, Provincia)

// DB2 Segreteria

STUDENTE(Matricola, Codice Fiscale, Cognome, Nome, Telefono, CFU Superati)

ESAMI2022(Codice Corso, Studente, Data, Voto, Lode)

- a) Integrare i due schemi tramite una vista ed utilizzando l'approccio GAV.
- b) Riferendosi alla soluzione di a), integrare lo schema globale tramite una vista ed utilizzando un approccio LAV (esamiLode/esamiSenzaLode) (studentiAV/studentiSA).

SOLUZIONE:

a)

```
Create view Studenti as (Select DB1.Studente.Matricola,  
                                .Cognome,  
                                .Nome,  
                                .Indirizzo,  
                                .CAP,  
                                .Città,  
                                .Provincia,  
                                DB2.Studente.Codice Fiscale,  
                                .telefono,  
                                .CFU  
From DB1.Studente  
      JOIN  
      DB2.Studente  
      ON  
      DB1.studente.matricola =  
      DB2.studente.matricola  
)
```

```
Create view Esami as (Select * From DB1.Esame2021  
                        UNION  
                        DB2.Esami2022  
)
```

b)

```
Create view EsamiConLode as (Select * From DB.Esami  
                             Where lode ="SI"  
                             )
```

```
Create view EsamiSenzaLode as (Select * From DB.Esami  
                               Where lode ="NO"  
                               )
```

```
Create view StudentiAV as (Select * From DB.Studenti  
                            Where provincia ="AV"  
                            )
```

```
Create view StudentiSA as (Select * From DB.Studenti  
                            Where provincia ="SA"  
                            )
```

*Assumendo che nel nostro DB abbiamo solo studenti di AV e di SA.

Esercizio

Data una tabella con 6 tuple. Si supponga che per una coppia di attributi X ed Y valgano le seguenti relazioni:

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

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

Calcolare il g3 error per una RFD $X \rightarrow Y$ che rilassi sull'extent.

SOLUZIONE:

costruire un esempio pratico può aiutarvi

- a) Togliendo t_1 e t_6 vale una FD canonica. Ciò vuol dire che esiste una RFD rilassata sul extent del tipo

$$D_{\text{TRUE}} : X_{\text{EQ}} \xrightarrow{\psi(X,Y) \leq 2/6} Y_{\text{EQ}}$$

	X	Y
t_1	A	1
t_2	A	2
t_3	A	2
t_4	B	2
t_5	B	2
t_6	B	3

Esercizio

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 .

- 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.
- Modificare le relazioni di cui sopra in modo che possa valere sia una FD che una RFD

SOLUZIONE:

costruire un esempio pratico può aiutarvi

- Non esiste una FD $X \rightarrow Y$, ma vale la seguente AFD:

$$D_{\text{TRUE}} : X_{\text{EQ}} \xrightarrow{\psi(X,Y) \leq 1/6} Y_{\text{EQ}}$$

	X	Y
t ₁	1	1
t ₂	1	1
t ₃	5	5
t ₄	6	6
t ₅	2	2
t ₆	1.5	2

-

b)

	X	Y
t ₁	1	1
t ₂	1	1
t ₃	5	5
t ₄	6	6
t ₅	2	2
t ₆	1.5	2

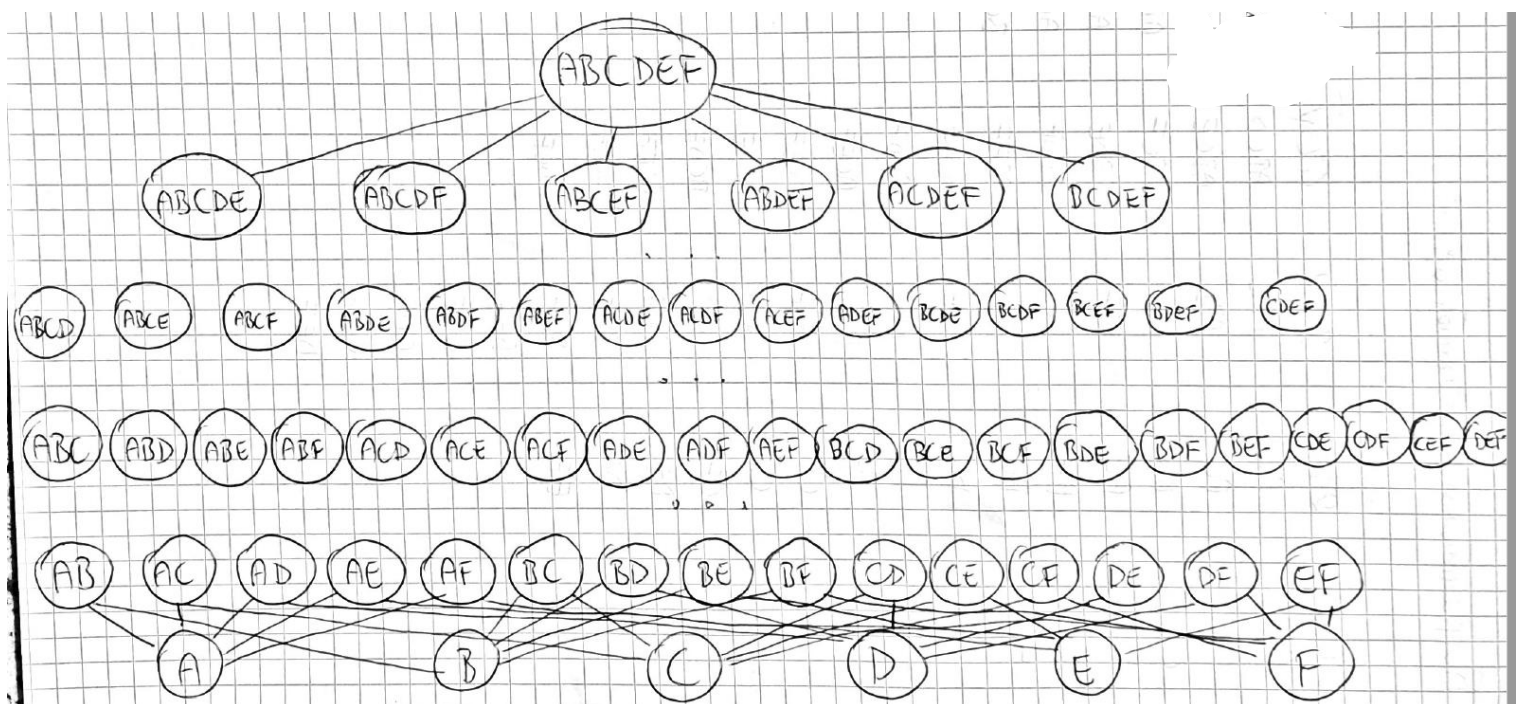
$$D_{\text{TRUE}} : X \xrightarrow[\max_{\Delta \in \Pi_X} \Delta \phi(\Delta[X]) \leq 0.5}]{\psi_{\text{gen}}(\phi)} Y \quad \max_{\Delta \in \Pi_X} \Delta \phi(\Delta[Y]) \leq 1$$

Esercizio

Disegnare la struttura a lattice per l'estrazione di dipendenze funzionali da un dataset con 6 attributi, indicando il numero totale di dipendenze candidate.

SOLUZIONE:

Ogni nodo del livello inferiore è collegato ai nodi del livello superiore che lo contengono



Rappresentazione alternativa:

1° LIV	2° LIV	3° LIV	4° LIV	5° LIV	6° LIV
A	AB	ABC	ABCD	ABCDE	ABCDEF
B	AC	ABD	ABCE	ABCDF	
C	AD	ABE	ABCF	ABCEF	
D	AE	ABF	ABDE	ABDEF	
E	AF	ACD	ABDF	ACDEF	
F	BC	ACE	ABEF	BCDEF	
	BD	ACF	ACDE		
	BE	ADE	ACDF		
	BF	ADF	ACEF		
	CD	AEF	ADEF		
	CE	BCD	BCDE		
	CF	BCE	BCDF		
	DE	BCF	BCEF		
	DF	BDE	BDEF		
	EF	BDF	CDEF		
		BEF			
		CDE			
		CDF			
		CEF			
		DEF			

Numero di nodi per livello

$$\text{FORMULA: } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

n = # di nodi k = livello

$$1^\circ \text{livello} = \binom{6}{1} = 6$$

$$4^\circ \text{livello} = 15$$

$$2^\circ \text{livello} = 15$$

$$5^\circ \text{livello} = 6$$

$$2^\circ \text{livello} = 20$$

$$6^\circ \text{livello} = 1$$

Numero di FD candidate

$$1^\circ \text{livello} = 30 \quad (15 \text{ nodi} * 2)$$

di archi tra 1° e 2° livello

$$2^\circ \text{livello} = 60 \quad (20 \text{ nodi} * 3)$$

di archi tra 2° e 3° livello

$$3^\circ \text{livello} = 60 \quad (15 \text{ nodi} * 4)$$

di archi tra 3° e 4° livello

$$4^\circ \text{livello} = 30 \quad (6 \text{ nodi} * 5)$$

di archi tra 4° e 5° livello

$$5^\circ \text{livello} = 6 \quad (1 \text{ nodi} * 6)$$

di archi tra 5° e 6° livello

$$6^\circ \text{livello} = 0$$