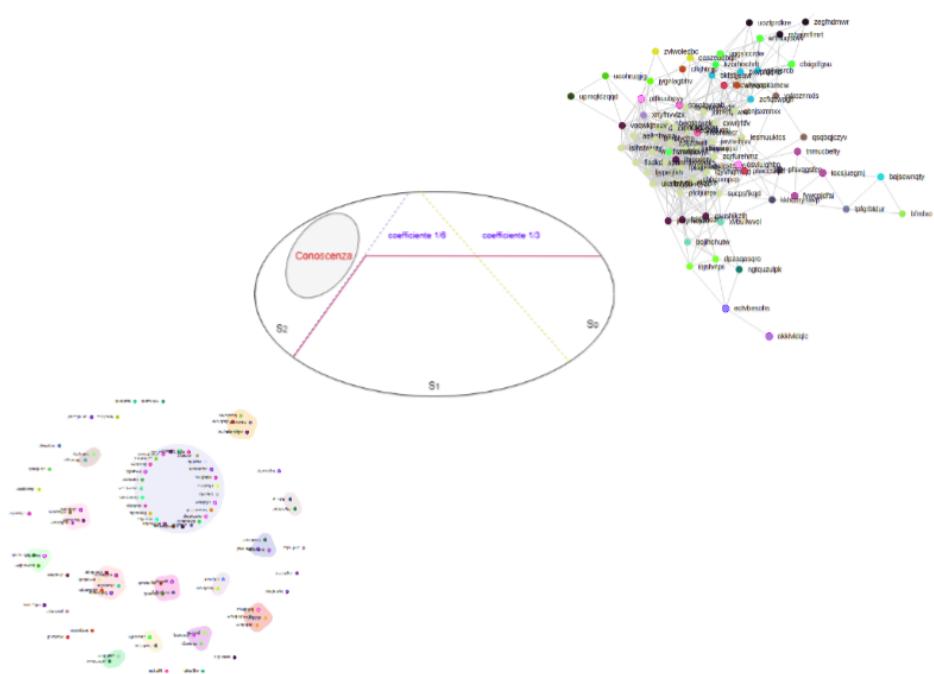


# Resoconto Tecnico

Documento contenente le attività ed i risultati ottenuti durante la generazione del Reticolo della Conoscenza, con riferimento alle risposte dei candidati durante colloquio di lavoro



# Indice

<b>1 Analisi dei dati di probabilità</b>	<b>4</b>
1.1 Problema in esame . . . . .	4
1.2 Caratteristiche degli eventi di coppia . . . . .	4
1.2.1 Eventi indipendenti . . . . .	5
1.2.2 Eventi dipendenti . . . . .	5
1.3 Evento conosciuto ed evento indovinato . . . . .	6
1.3.1 Probabilità di rispondere correttamente ad una domanda . . . . .	7
1.3.2 Il piano . . . . .	9
<b>2 Rete neurale</b>	<b>14</b>
2.1 Test effettuati . . . . .	15
2.1.1 Architettura della rete: 4 neuroni per ciascuno dei 2 layers . . . . .	16
2.1.2 Architettura della rete a 2 neuroni per ciascuno dei 2 layers . . . . .	17
2.1.3 Architettura della rete a 4 neuroni per 1 layer . . . . .	20
2.2 Sviluppo della rete delle domande nel database . . . . .	21
2.2.1 Montaggio e configurazione della rete per le domande di logica . . . . .	21
2.2.2 Test e Documentazione . . . . .	23
2.3 Funzionalità offerte dall'interfaccia della Rete neurale costruita	41
2.3.1 Diagrammi dei casi d'uso . . . . .	42
2.3.2 Requisiti . . . . .	68
2.3.3 Interfaccia utente Rete neurale - Definitiva . . . . .	73
<b>3 Principal Component Analysis</b>	<b>80</b>
3.1 Metodologia applicata . . . . .	80
3.2 Sviluppo . . . . .	81
3.3 Risultati ottenuti . . . . .	83
3.3.1 Calcolo della Principal Component Analysis . . . . .	83
3.3.2 Calcolo degli autovettori ed individuazione di Summary	84
3.3.3 Calcolo degli autovettori . . . . .	92
3.3.4 Calcolo della matrice di correlazione . . . . .	92
3.4 Conclusione dell'analisi . . . . .	95
<b>4 Costruzione del Reticolo della Conoscenza</b>	<b>98</b>
4.1 Descrizione del sistema . . . . .	98
4.1.1 Configurazione . . . . .	99
4.2 Creazione dei file CSV . . . . .	100
4.3 Creazione del Reticolo della Conoscenza per sui dati di Prova	100
4.3.1 L'architettura della Rete . . . . .	100

4.3.2	Reticolo della Conoscenza generato mediante previsioni sul set di dati che conto esclusivamente della conoscenza del candidato . . . . .	101
4.3.3	Reticolo della Conoscenza generato mediante la frequenza calcolata sul set di dati che tiene conto esclusivamente della conoscenza del candidato . . . . .	107
4.3.4	Osservazioni sul Reticolo generato dalle previsioni della rete . . . . .	109
4.3.5	Reticolo della Conoscenza generato con set di dati che tiene conto della possibilità che un candidato ha di indovinare la risposta . . . . .	110
4.4	Creazione del Reticolo della Conoscenza per le domande di logica nel database . . . . .	111
4.4.1	Individuazione dei gruppi di argomenti . . . . .	112
4.4.2	Architettura della Rete testata . . . . .	114
4.4.3	Risultati di frequenza sulle domande di logica . . . . .	124
4.5	Creazione del Reticolo della Conoscenza per delle domande di SQL nel database . . . . .	125
4.5.1	Individuazione dei gruppi di argomenti . . . . .	125
4.5.2	Architettura della Rete testata . . . . .	128
4.5.3	Risultati di frequenza sulle domande SQL . . . . .	133
<b>5</b>	<b>Configurazione dell'applicativo progettato e strumenti di sviluppo</b>	<b>135</b>
5.1	Test d'unità . . . . .	135
5.2	Sistema operativo e browser . . . . .	136
5.2.1	Specifiche di sviluppo . . . . .	136

# 1 Analisi dei dati di probabilità

Durante il periodo 20/05 - 24/05 mi sono occupata di analizzare la probabilità che ha un candidato di rispondere correttamente alle domande in fase di test; valutando le relazioni di dipendenza che possono esistere tra più domande e l'impatto che può assumere la fortuna.

## 1.1 Problema in esame

Test, sottoposto ad un candidato durante un colloquio, composto da *domande a tripla risposta multipla*.

Nel suddetta sezione vengono analizzate le relazioni che intercorrono tra due domande, denominate A e B, a seconda se il candidato risulta in grado di rispondervi correttamente o meno.

## 1.2 Caratteristiche degli eventi di coppia

Tipi di eventi trattati:

- **Eventi indipendenti;**
- **Eventi dipendenti:**
  - A e B sono strettamente dipendenti;
  - A implica B.
- **Evento conosciuto ed evento indovinato.**

Struttura usata per rappresentare la probabilità degli eventi di coppia:

$$\begin{array}{c} AB \\ /\backslash \\ A \ B \\ \backslash / \\ Z \end{array}$$

con:

- $AB$  rappresenta la probabilità complessiva dell'evento che si verifica sempre;
- $A$  rappresenta la probabilità che permette il verificarsi di A, ma non di B;

- $B$  rappresenta la probabilità che permette il verificarsi di  $B$ , ma non di  $A$ ;
- $Z$  rappresenta la probabilità a zero, l'impossibilità del verificarsi dell'evento.

### 1.2.1 Eventi indipendenti

$A$  e  $B$  sono due domande la quali risposte sono completamente scorrelate tra di loro.

$$\begin{array}{c}
 P(A)P(B) \\
 /\backslash \\
 P(A)(1 - P(B)) \ (1 - P(A))P(B) \\
 \backslash / \\
 (1 - P(A))(1 - P(B))
 \end{array}$$

**Considerazioni generali** La probabilità complessiva nel caso di domande indipendenti  $A$  e  $B$  viene data da  $P(A)$  per  $P(B)$ .

Se è conosciuta dal candidato la risposta alla domanda  $A$  ma non alla domanda  $B$  la probabilità di ottenere una risposta corretta è  $P(A)$ , mentre la probabilità di ottenere una risposta non corretta per  $B$  vale  $1 - P(B)$ . Il ragionamento duale è svolto nel calcolo della probabilità per la risposta corretta alla domanda  $B$  ma non ad  $A$ .

La probabilità di non ottenere alcuna risposta corretta alle due domande viene calcolata prendendo in considerazione gli eventi contrari a quelli coinvolti. Dunque per  $A$  la probabilità che il candidato non conosca la soluzione è  $1 - P(B)$ , dualmente per  $B$  la probabilità è  $1 - P(A)$ .

### 1.2.2 Eventi dipendenti

**A e B sono due domande fortemente correlate tra di loro** se si risponde correttamente ad una delle due domande si risponde correttamente anche all'altra.

$$\begin{array}{c}
 P(A)^2 \\
 /\backslash \\
 0\ 0 \\
 \backslash / \\
 (1 - P(A))^2
 \end{array}$$

**Considerazioni generali** La probabilità complessiva nel caso di domande dipendenti A e B viene data da  $P(A)$  per  $P(B)$ ; ma  $P(A) = P(B)$  dunque  $P(A)^2 = P(B)^2$ .

Conseguentemente se il candidato non conosce la risposta alla domanda A non può conoscere la risposta alla domanda B perciò la probabilità di conoscere uno dei due eventi è pari a 0.

In questo caso la probabilità a 0 è  $(1 - P(A))(1 - P(B)) = (1 - P(A))^2$  essendo che A=B.

**A implica B** Se si sa rispondere alla domanda A di conseguenza si è in grado di rispondere anche alla domanda B.

Tuttavia non vale il ragionamento opposto, se si sa rispondere alla domanda B non significa che si è in grado di rispondere alla domanda A.

$$\begin{array}{c} P(A) \\ /\backslash \\ 0 \quad P(B) - P(A) \\ \backslash / \\ 1 - P(B) \end{array}$$

**Considerazioni generali** La probabilità complessiva nel caso di domande dipendenti A e B viene data esclusivamente da  $P(A)$ , in quanto la conoscenza di sia di A che di B è possibile solo se si ha piena conoscenza di A.

Dunque la probabilità che si conosca la risposta alla domanda A ma non a B è impossibile (pari a 0); mentre se si ha conoscenza della domanda B ma non di A la probabilità si stanzia a  $P(B) - P(A)$ .

La probabilità a zero è  $1 - P(B)$  indicatore dell'impossibilità di avere la risposta corretta per A.

### 1.3 Evento conosciuto ed evento indovinato

Durante un test il candidato deve saper scegliere la risposta, corretta o meno, alla domanda posta. Le variabili che entrano in gioco durante l'esecuzione dell'atto non riguardano esclusivamente la conoscenza personale del singolo. La probabilità di un evento A è data dalla formula:

$$P(A) = P(A_C) + P(A_I)$$

Le variabili in uso sono:

- $P(A_C)$ : probabilità che il candidato sappia rispondere alla domanda A correttamente per sua conoscenza;
- $P(A_I)$ : probabilità che il candidato sappia rispondere alla domanda A correttamente indovinando.

Per quanto appena definito sopra valgono le seguenti proprietà:

1.  $P(B_C|A_C) = 1$
2.  $P(B_C|A_I) = P(B_C)$
3.  $P(B_I|A_C) = 0$
4.  $P(B_I|A_I) = P(B_I)$

### 1.3.1 Probabilità di rispondere correttamente ad una domanda

Variabili coinvolte:

- $P(A)$ : probabilità necessaria perchè si verifichi, per la domanda A, che il candidato dia la risposta corretta. Per la legge dei grandi numeri la frequenza porta alla probabilità.
- $S_0$ : insieme dei casi in cui in una domanda non viene scartata alcuna risposta dal dominio delle risposte possibili;
- $S_1$ : insieme dei casi in cui in una domanda viene scartata una risposta dal dominio delle risposte possibili;
- $S_2$ : insieme dei casi in cui in una domanda vengono scartate due risposte dal dominio delle risposte possibili.
- $P(I)$ : probabilità di dare la risposta corretta alla domanda A indovinando;
- $P(C)$ : probabilità di dare la risposta corretta alla domanda A per conoscenza.

Sapendo che  $P(I) = P(A) - P(C)$  logicamente vale anche  $P(A) = P(I) + P(C)$ .

Se un candidato non è in grado scartare alcuna risposta dalla domanda ha 1 possibilità su 3 di, indovinando, dare la risposta corretta. Se un candidato invece risulta in grado di scartare una risposta sbagliata alla domanda, rimane con 1 possibilità su 2 di poter dare la risposta corretta. Se invece, caso ottimo, il candidato ha piena conoscenza della domanda posta risulta in grado di scartare due risposte sbagliate lasciando un'unica risposta possibile, quella esatta. Il ragionamento sopra espresso può venire espresso con la seguente espressione:

$$P(A) = P(S_0)\frac{1}{3} + P(S_1)\frac{1}{2} + P(S_2)$$

Ora individuiamo quale è la probabilità effettiva per un candidato di dare la risposta corretta ad una domanda A.

$$\begin{aligned} 1 &= S_0 + S_1 + S_2 \\ S_0 &= 1 - S_1 - S_2 \end{aligned}$$

Sostituendo:

$$\begin{aligned} P(A) &= (1 - P(S_1) - P(S_2))\frac{1}{3} + P(S_1)\frac{1}{2} + P(S_2) \\ &= \frac{1}{3} - \frac{1}{3}P(S_1) - \frac{1}{3}P(S_2) + \frac{1}{2}P(S_1) + P(S_2) \\ &= \frac{1}{3} + \frac{1}{6}P(S_1) + \frac{2}{3}P(S_2) \end{aligned}$$

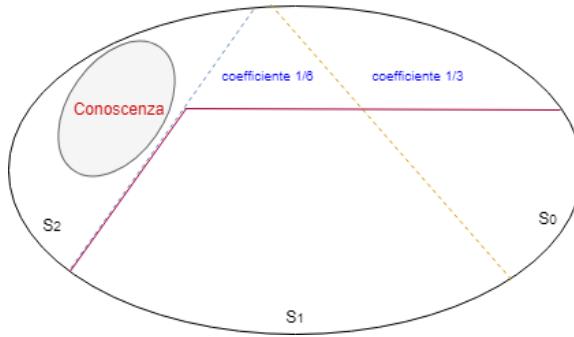


Figura 1: Rappresentazione insiemistica della probabilità di rispondere correttamente ad una domanda:  $P(A)$

### Considerazioni importanti

In conclusione  $P(A) = \frac{1}{3} + \frac{1}{6}P(S_1) + \frac{2}{3}P(S_2)$ . Ovvero la probabilità per un candidato di dare in una domanda A la risposta corretta dipende dai seguenti fattori:

- $\frac{1}{3}$ : coefficiente che rappresenta la probabilità effettiva per chi non conosce la risposta alla domanda di dare la risposta corretta;
- $\frac{1}{2}P(S_1)$ : coefficiente che rappresenta la probabilità effettiva di dare la risposta corretta quando il candidato è in grado di scartare una risposta sbagliata alla domanda;
- $\frac{2}{3}P(S_2)$ : coefficiente che rappresenta la probabilità effettiva di dare la risposta corretta quando il candidato è in grado di scartare due risposte sbagliate alla domanda.

Dall’analisi della tipologia di eventi di coppia e dal calcolo della probabilità necessaria per poter rispondere correttamente ad una domanda, si giunge alla valenza dei seguenti assiomi:

1. Le coppie di domande A e B devono essere fra loro disgiunte, altrimenti si genererebbero situazioni di invalidità dei risultati;
2. Per rispondere correttamente ad una domanda non è necessario che il candidato abbia piena conoscenza di tutti gli argomenti richiesti dall’esame, ma bensì ne risultano sufficienti  $n - 1$ ;
3. La probabilità di conoscere è contenuta all’interno di  $S_2$ , in quanto se un candidato conosce è conseguentemente in grado, da una domanda, di scartare due risposte sbagliate.

### 1.3.2 Il piano

La probabilità  $P(A)$  che un candidato ha in gioco nel momento in cui si approccia a rispondere ad una domanda può venire rappresentata in un piano.

Ognuno dei tre assi cartesiani rappresenta l’insieme dei casi di scarto ( $S_0$ ,  $S_1$ ,  $S_2$ ). L’intersezione tra i punti del piano indica la regione accettabile contenente il range di valori assumibili da  $P(A)$ . Tale punto proiettato su ognuno dei tre assi permette l’individuazione esatta dei coefficienti delle variabili  $S_0$ ,  $S_1$ ,  $S_2$ .

Ogni porzione della regione del piano viene individuata con la seguente tecnica:

1. Per individuare ogni retta passante per  $S_0$ ,  $S_1$  e  $S_2$  è necessario assumere che  $S_0 + S_1 + S_2 = 1$ ;
2. La retta passante per  $S_0$  è rappresentabile per mezzo delle seguenti equazioni:

$$S_0 = 0 \text{ e } S_1 + S_2 = 1$$

In questo modo l’asse  $S_0$  è fissato a 0 e estrapolando  $S_1$  e  $S_2$  da  $S_1 = -S_2 + 1$  assumono valori tra (0,1).

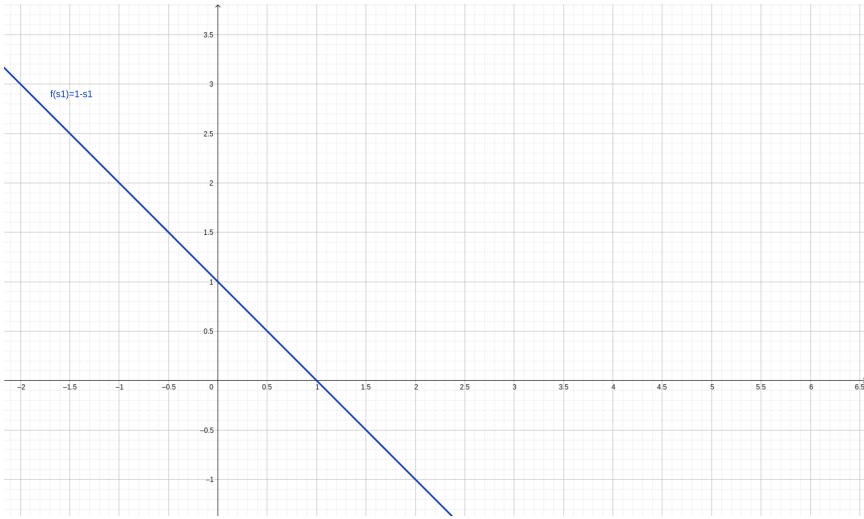


Figura 2: Rappresentazione della retta passante per  $S_0 = 0$

3. Il medesimo ragionamento vale per le rette passanti per  $S_1$  e  $S_2$ .

$$S_1 = 0 \text{ e } S_0 + S_2 = 1$$

l'asse  $S_1$  è fissato a 0 e  $S_0$  e  $S_2$  assumono valori tra (0,1).

$$S_2 = 0 \text{ e } S_1 + S_0 = 1$$

l'asse  $S_2$  è fissato a 0 e  $S_0$  e  $S_2$  assumono valori tra (0,1).

4. In questo modo l'unione di tutte le rette passanti per gli assi creano la regione accettabile dei valori di  $P(A)$ .

Avendo rappresentato il piano, si ottiene nei punti di intersezioni fra le tre rette, la regione accettabile per  $P(A)$ . Inoltre è possibile individuare il fascio di rette, che tangenti il piano, permettono di affermare se una specifica domanda è, in base alla sua frequenza, di difficoltà bassa, media, alta per un candidato.

- Se una domanda ha una difficoltà bassa la retta si situa passante per i punti  $0 < S_2 \leq 1$  (molto vicino a 1) e  $(S_0, S_1) < 0$  (tendenti a 0);
- Se una domanda ha una difficoltà alta la retta si situa passante per i punti  $S_2 \leq 0$  (molto vicino a 0),  $S_1 < 1$  e  $S_0 \leq 1$  (tendente a non scartare alcuna risposta);

- Se una domanda ha una difficoltà media la retta si situa nella parte centrale della regione accettabile, passante per i punti  $0 \leq (S_0, S_1, S_2) \leq 1$ .

### Rappresentazione di $P(A)$

Vediamo alcuni casi di come le domande possono venire rappresentate sul piano:

La funzione di partenza è:

$$F = \frac{1}{3} + \frac{1}{6}S_1 + \frac{2}{3}S_2$$

Va esplicitato  $S_1$ , i passaggi utili da fare sono i seguenti:

$$\frac{-1}{6}S_1 = \frac{1}{3} + \frac{2}{3}S_2 - F \rightarrow S_1 = -4S_2 - 2 + 6F$$

Essendo che  $0 \leq S_2 \leq 1$  usando  $S_1 = 1$  e  $S_2 = 0$  si ottiene che  $F = \frac{1}{2} = 0.5$

Quanto appena calcolato può venire rappresentato graficamente impiegando la retta  $S_1 = 1 - S_2$  (responsabile di definire una porzione del piano in base alle variaibili coinvolte) e mediante la retta  $S_1 = -4S_2 - 2 + 6F$  (che permette di calcolare il fascio di rette tangenti alla prima retta).



Figura 3: Rappresentazione di  $P(A)$  per una frequenza 0.5 proiettata su assi  $S_0 = 0$ ,  $S_1$  e  $S_2$ .

Nella figura sopra sono rappresentati i seguenti significati:

- La linea azzurra rappresenta  $S_2 = 1 - S_1$ ;
- La linea rosa rappresenta la retta tangente  $S_1 = -4S_2 + 1$ ;
- Punto A (blu):

$$\begin{aligned} S_1 &= 0.5 = \frac{1}{2} \\ \frac{1}{2} &= -4S_2 + 1 \rightarrow 4S_2 = 1 - \frac{1}{2} \rightarrow S_2 = \frac{1}{8} \\ S_0 &= 1 - \frac{1}{2} - \frac{1}{8} = \frac{3}{8} \end{aligned}$$

Ovvero metà dei candidati sottoposti alla domanda sa scartare una delle risposte, lo 0.16% sa dare la risposta corretta e lo 0.36% non sa scartare alcune delle risposte possibili.

- Punto B (verde):

$$\begin{aligned} S_1 &= 0 \\ S_2 &= \frac{1}{4} \\ S_0 &= 1 - \frac{1}{4} = \frac{3}{4} \end{aligned}$$

Ovvero nessun dei candidati sottoposti alla domanda sa scartare una delle risposte, lo 0.25% sa dare la risposta corretta e lo 0.75% non sa scartare alcune delle risposte possibili.

- Punto C (fucsia):

$$\begin{aligned} S_1 &= \frac{1}{4} \\ S_2 &= \frac{3}{16} \\ S_0 &= 1 - \frac{1}{4} - \frac{3}{16} = \frac{9}{16} \end{aligned}$$

Ovvero lo 0.25% dei candidati sottoposti alla domanda sa scartare una delle risposte, lo 0.19% sa dare la risposta corretta e lo 0.56% non sa scartare alcune delle risposte possibili.

- Punto D (arancione):

$$\begin{aligned} S_1 &= \frac{1}{3} \\ S_2 &= \frac{1}{3} \\ S_0 &= 1 - \frac{1}{3} - \frac{1}{3} = \frac{1}{3} \end{aligned}$$

Osserviamo che il punto in esame fuoriesce dalla regione delimitata dalla retta tangente di frequenza 0.5 ( $S_1 = -4S_2 + 1$ ). Conseguenza diretta data dall'impossibilità di ottenere una probabilità del 50% sulla domanda con  $\frac{1}{3}$  di candidati che sa scartare 2 risposte,  $\frac{1}{3}$  che ne sa scartare 1 e  $\frac{1}{3}$  nessuna.

Vediamo ulteriori due esempi che permettono di valutare cosa accade nel piano nel caso di una frequenza:

1. Quasi in prossimità di 1;

2. Pari alla soglia minima dell'indovinato.

Il grafico è il seguente:



Figura 4: Rappresentazione di  $P(A)$  per una frequenza 0.33 e 0.8 proiettate su assi  $S_0 = 0$ ,  $S_1$  e  $S_2$ .

- La linea grigia mostra la retta tangente con frequenza 0.80%. In questa abbiamo calcolato il punto E (giallo):

$$\begin{aligned} S_1 &= 0 \\ S_2 &= \frac{14}{20} \\ S_0 &= 0 \end{aligned}$$

Quasi la totalità dei candidati ha la conoscenza per poter scartare tutte le risposte sbagliate e dare la risposta giusta alla domanda.

- La linea viola mostra la retta tangente con frequenza 0.33%. In questa abbiamo calcolato il punto F (rosa):

$$\begin{aligned} S_1 &= 0 \\ S_2 &= 0 \\ S_0 &= 1 \end{aligned}$$

Ovvero nessuno dei candidati ha la conoscenza per poter scartare né una nè due risposte, percui l'unica possibilità per un candidato di rispondere alla domanda è indovinare. È evidente come se un candidato non sa la risposta ad una domanda ha una probabilità dello 0.33% di poter indovinare la risposta corretta.

## 2 Rete neurale

La libreria utilizzata per sviluppare la Rete neurale è stata *ConvNetJS*. L'aspetto positivo di tale scelta è la semplicità nell'utilizzo del linguaggio javascript e la possibilità di implementare funzioni complesse dotate di interfaccia grafica. L'aspetto negativo riguarda la totale mancanza di mantenibilità della libreria, con scarsa presenza di esempi applicativi oltre alla documentazione ufficiale, che costringono lo sviluppatore ad una ricerca approfondita personale e a continui test per validare i risultati prodotti.

### Questions test - Revision Neural Net

**CONFIGURAZIONE RETE NEURALE**

```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:6});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:6});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01, momentum:0.1, batch_size:10, l2_decay:0.001});
```

**OUTPUT DELLA RETE**

```
Richiesta di previsione inviata alla rete ...
Richiesto di previsione accettata
Il vettore [0,1,0,0,0,0] ha previsione calcolata di [-0.14178025279470646,-0.014639237878831693,-0.08131295482446743,0.11790470015357472,0.14410273493990078,0.1329311955165462]
Rete Neurale in attesa ... Inserire risultati del test...
Inserire risultati del test... Nel test andato a buon fine
Ricapitolazione dati inseriti: Risposte ottenute [0,-1,0,0,0,0]
Inizio allenamento della rete ...
Richiesta di previsione inviata alla rete ...
Richiesto di previsione accettata
Il vettore [0,1,1,0,0,0] ha previsione calcolata di [-0.9703841414997074,-0.7824502014273779,0.34740891066895563,-0.8448991676224333,-0.5825080715009475,0.2936985044998004]
Rete Neurale in attesa ...
```

**INSERIMENTO DATI DI ADDESTRAMENTO**

Number of fields (max 9999): 6

Input1: 0  Input2: -1  Input3: 0  Input4: 0  Input5: 0  Input6: 0

Vettore di previsione:

Figura 5: Primo prototipo di interfaccia utente della Rete neurale di prova.

Durante il periodo 24/05 - 31/05 mi sono occupata dello sviluppo di una Rete neurale in grado di ricevere in input un training set di dimensione 6 e di restituire una previsione sui dati di apprendimento generati. Il problema che la rete mira ad analizzare è quello discusso nel capitolo §1.

Per agevolare l'apprendimento della rete, ed ottenere delle previsioni stabili, ho implementato due metodi di generazione randomica di dati, in modo da far apprendere massicciamente la stessa. Il dato prodotto consiste in un vettore di 6 elementi, composto da -1, 0 e 1 con il seguente criterio:

- **-1:** la domanda  $x$  è stata posta al candidato che ha risposto in maniera errata;

- **0:** la domanda x non è stata posta al candidato;
- **1:** la domanda x è stata posta al candidato che ha saputo rispondere correttamente.

Il primo metodo sviluppato si occupa di generare un vettore di dati di apprendimento basandosi esclusivamente su come le domande sono interconnesse tra di loro (grazie all'uso di un Grafo della Conoscenza costruito ad hoc). Il secondo metodo ripropone quanto perseguito dal primo, con il valore aggiunto, di generazione di un profilo randomico per un candidato, che tiene conto della probabilità di risposta ad una domande seguendo la formula  $P(A) = \frac{1}{3} + \frac{1}{6}P(S_1) + \frac{2}{3}P(S_2)$ .

## 2.1 Test effettuati

Quando si parla di domande nel database, si fa riferimento principalmente ai test di logica svolti dai candidati.

Alcune decisioni che ho preso per individuare l'architettura adeguata di rete possono essere così riassunte:

1. Una rete neurale non deve, per fornire dei dati attendibili, possedere un numero di neuroni troppo elevato rispetto al trainset effettuato; altrimenti la previsione ritornerebbe l'identità del vettore di input della stessa, come conseguenza diretta della capacità troppo elevata di immagazzinare dati.
2. I layers, ho deciso, di allenarli mediante tecnica di regressione, che permette l'inserimento in input di una funzione obiettivo e l'ottenimento di un risultato. L'output, anche in virgola mobile è composto da tanti elementi quanti sono i neuroni di regressione dichiarati. Per la mia rete di prova è sufficiente dichiarare 6 neuroni in regressione perchè l'output, appunto, che ci si aspetta dal sistema è di 6 elementi.
3. Per costruire un dataset di dati consistente, che consenta alla rete di imparare qualcosa, ho costruito un Grafo della Conoscenza con lo scopo di mettere in relazione degli argomenti d'interesse di più domande.

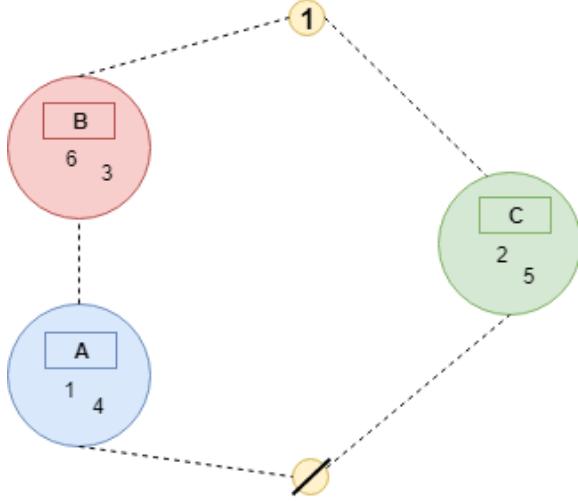


Figura 6: Grafo della Conoscenza rappresentante le relazioni esistenti tra il set di domande di prova.

Per svolgere l'apprendimento ogni vettore, facente parte del dataset, viene dato in pasto alla rete, che a sua volta provvede alla sua assimilazione, conseguenza diretta della tecnica dell'autoencoder, ovvero la rete impara il vettore riducendone lo spazio occupato.

4. Per creare il dataset ho ritenuto sufficiente generare *2000* vettori di risposta in modo da compiere in maniera esaustivo l'apprendimento della rete.

La tematica della validità dell'architettura viene approfondita nel dettaglio della sezione 4.3.1.

Il vettore passato in input per svolgere le previsioni è  $[0,0,0,0,0,0]$ ,  $[0,0,1,0,1,0]$  e  $[0,0,-1,0,0,0]$

Le aspettative riguardano la previsione di risposta di un candidato . Di seguito riporto quanto è stato rilevato in fase di test.

### 2.1.1 Architettura della rete: 4 neuroni per ciascuno dei 2 layers

Architettura della rete utilizzata:

```

layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:6});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:6});

net = new convnetjs.Net();

```

```

net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```

I layers utilizzati sono 2 e composti da 4 neuroni.

### **Training set standard a 4 neuroni per ciascuno dei 2 layers**

- Il vettore [0,0,0,0,0,0] ha previsione calcolata di  
[-0.021598804903572744, -0.1372509042342871, 0.06611969158456255,  
0.018121335417653706, -0.11264571886853292, 0.17520370837747462]

Appaiono in relazione le domande 1, 2, 5 e 3, 4, 6.

Gli scostamenti tra le coppie 2, 5 e 3, 6 sono consistenti con quelle che sono le relazioni di dipendenza; invece 1, 4 ha una differenza di 0.016 circa che parte da qualche millesimo fino 0.5. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4; nel test in analisi questo non viene rispettato da nessuna delle coppie in analisi per differenze che vanno da qualche millesimo fino a 0.018 circa.

### **Osservazioni**

L'architettura testata si compone di 4 neuroni a layer su una base di 2000 test, con il rischio di avere una rete che apprende troppo e che come effetto negativo "veda" addirittura cose che non esistono. A prova di ciò sono i risultati non conformi alle attese. Ho scelto di conseguenza, di ridurre il numero di neuroni presenti in ciascun layers e/o il numero di layers presenti.

#### **2.1.2 Architettura della rete a 2 neuroni per ciascuno dei 2 layers**

Architettura della rete utilizzata:

```

layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:6});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:6});

net = new convnetjs.Net();

```

```

net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```

I layers utilizzati sono 2 composti da 2 neuroni.

### **Training set standard su rete a 2 neuroni per ciascuno dei 2 layers**

- Il vettore [0,0,0,0,0,0] ha previsione calcolata di [0.31232372051574936, 0.7253754889487585, -0.5051208979797573, 0.32075742158673093, 0.7324947496336937, -0.4348299972940168] Appaiono in relazione le domande 1, 2, 4, 5 e 3, 6. Gli scostamenti tra le coppie 1, 4 e 3, 6 e 2, 5 sono consistenti con quelle che sono le relazioni di dipendenza fra le domande. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4; in questo test la regola viene rispettata pienamente. Dai dati della previsione si nota come il candidato ha una buona probabilità di saper rispondere alla coppia 1 e 4, e ancora più elevata di saper rispondere correttamente alla coppia 2 e 5; molto bassa di saper rispondere correttamente alle 3 e 6 che sono, appunto, di una difficoltà maggiore rispetto alla coppia 1 e 4.
- Il vettore [0,0,1,0,1,0] ha previsione calcolata di [0.5123144717131076, 0.9123354449531641, 0.2837937822420923, 0.46449868699771607, 0.9029832167165894, 0.3227303792035435] Appaiono in relazione le domande 1, 2, 3 4, 5, 6. Gli scostamenti tra le coppie 1, 4 e 3, 6 e 2, 5 sono consistenti con quelle che sono le relazioni di dipendenza fra le domande. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4; in questo test la regola viene rispettata pienamente. Dai dati della previsione si nota come il candidato ha un'ottima probabilità di saper rispondere alla coppia 2 e 5 (come imposto dal vettore previsione), buona di saper rispondere alla coppie 3 e 6 (come imposto dal vettore previsione) e più che buona di saper rispondere alle 1 e 4, che sono di una semplicità più elevata rispetto alla 3 e 4.
- Il vettore [0,0,-1,0,0,0] ha previsione calcolata di [0.3698539826215957, 0.288907514487717, -0.8504159455662308, 0.3663192502433841, 0.2937448801761998, -0.7845589473185985]

Appaiono in relazione le domande 1, 2, 4, 5 e 3, 6.

Gli scostamenti tra le coppie 1, 4 e 3, 6 e 2, 5 sono consistenti con quelle che sono le relazioni di dipendenza fra le domande. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4; in questo test la regola viene rispettata pienamente.

Dai dati della previsione si nota come il candidato ha una discreta probabilità di saper rispondere alla coppia 2 e 5, un pò meglio di saper rispondere alla coppia 1 e 4 e di non saper saper rispondere alle 3 e 6 (come imposto dal vettore previsione).

#### **Training set con generazione del profilo di un candidato con probabilità su rete a 2 neuroni per ciascuno dei 2 layers**

- Il vettore  $[0,0,0,0,0,0]$  ha previsione calcolata di  
 $[0.057781303506280995, 0.0513731100126314, -0.06600467867066256,$   
 $0.029940883111932555, -0.019564515397168573, -0.09570617900597932]$

Appaiono in relazione le domande 1, 2, 4 e 3, 5, 6.

Gli scostamenti tra la coppia 1, 4 e 3, 6 sono consistenti con quelle che sono le relazioni di dipendenza fra le domande; invece per la coppia 2 e 5 i segni sono opposti con una differenza di 0.024. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4, la regola viene rispettata pienamente. Le anomalie riscontrate sono da ricondurre alla natura del vettore di training che si basa sul calcolo della probabilità di una risposta.

- Il vettore  $[0,0,1,0,1,0]$  ha previsione calcolata di  
 $[0.19494624113789977, 0.1712744021266377, 0.577963304906936,$   
 $0.781098215373483, 0.3774535909060714, 0.03617314870307162]$

Appaiono in relazione le domande 1, 2, 3, 4, 5, 6.

Gli scostamenti tra le coppie 1 e 4, 2 e 5, 3 e 6 sono consistenti con quelle che sono le relazioni di dipendenza fra le domande. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4, la regola non viene rispettata dalla domanda 1 in rapporto con la domanda per una differenza di 0.37 circa. Le anomalie riscontrate sono da ricondurre alla natura del vettore di training che si basa sul calcolo della probabilità di una risposta.

- Il vettore  $[0,0,-1,0,0,0]$  ha previsione calcolata di  
 $[0.09845785763965222, 0.015421380649956663, -0.5138068038427066,$   
 $-0.4853190165287735, -0.22629262719814794, 0.0008152164571250502]$

Appaiono in relazione le domande 1, 2, 6 e 3, 4, 5.

Gli scostamenti tra le coppie 1, 4 e 2, 5 e 3, 6 per una differenza tuttavia trascurabile che oscilla dallo 0.2 allo 0.5. Le domande 3 e 6

si dovrebbero presentare con una positività inferiore rispetto a 1 e 4, la regola non vale per la coppia 6 e 4. Le anomalie riscontrate sono da ricondurre alla natura del vettore di training che si basa sul calcolo della probabilità di una risposta.

### Osservazioni

Confrontando i risultati ottenuti dalla rete con i layers impostati a 4 neuroni con quanto emerso dai dati risultanti dalla rete a 2 neuroni; emerge come l'architettura a 2 neuroni a layers è sicuramente quella che da i risultati attesi.

Quanto osservato di discordate nel secondo training set è, come da aspettative, da associare alla natura stessa della creazione del set di dati.

#### 2.1.3 Architettura della rete a 4 neuroni per 1 layer

Architettura della rete utilizzata:

```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:6});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:6});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```

Viene utilizzato un unico layer da 4 neuroni.

#### Training set standard su rete a 4 neuroni per 1 layer

- Il vettore [0,0,0,0,0,0] ha previsione calcolata di [0.12202628618565468, 0.08221724740100582, 0.02233631914718809, 0.09586625658118901, 0.05558075220027264, 0.13443779128784109]

Appaiono in relazione le domande 1, 2, 3, 4, 5, 6.

Gli scostamenti tra le coppie 1, 4 e 3, 6 e 2, 5 sono consistenti con quelle che sono le relazioni di dipendenza fra le domande. Le domande 3 e 6 si dovrebbero presentare con una positività inferiore rispetto a 1 e 4; in questo test la regola non viene rispettata dalla domanda 6.

Dai dati della previsione si nota come il candidato non ha una buona probabilità di saper rispondere alle domande e la domanda 6 non si presenta conforme alle aspettative.

## Osservazioni

Rispetto a quanto osservato nei casi precedenti, ancora l'architettura che rispetta le attese è quella con 2 neuroni per 2 layers.

Tale conclusione ha perfettamente senso, in quanto il Grafo della Conoscenza, che ho usato come base per costruire i vettori di apprendimento, è composto da 3 nodi (A, B, C) indicanti 3 neuroni. Il quarto può venire valutato come un nodo della rete utile per parametri in entrata e in uscita.

Per estendere maggiormente la mia conoscenza della rete, ho provveduto ad aumentare progressivamente il numero di neuroni a layers e osservarne le interazioni. Svolgendo ciò mi sono accorta che il risultato ottenuto dalla previsione era il più possibile vicino al vettore previsione quando facevo uso di un numero eccessivo di neuroni rispetto al training set svolto. La situazione è generatrice di overfitting e non attendibilità dei dati raccolti. Un'architettura con 1 neurone nel primo e 2 nel secondo layer , invece, presenta una buona capacità di previsione in quasi tutti i casi, però tende ad andare in underfitting, come riporto di seguito:

```
Il vettore [0,0,0,0,0,1] ha previsione calcolata di  
[0.5613347853884025, 0.8310670629630683, -1.03049430206139,  
0.5492731069379962, 0.5679700877862532, -0.8637707232817535]
```

Il numero di neuroni non è sufficiente per memorizzare che la domanda 6 deve essere positiva, e comporta a cascata la non correttezza anche delle domande 3, 4 e 1. La situazione si presenta simile, se il layer con 1 neurone è posto al di sotto.

## 2.2 Sviluppo della rete delle domande nel database

### 2.2.1 Montaggio e configurazione della rete per le domande di logica

Durante la settimana dal 03/06 al 07/06 la mia attività principale è stata il montaggio e la configurazione della Rete neurale, inerente il database aziendale con dataset i colloqui ai candidati per le domande di logica. Inoltre ho rivolto parte delle ore a modificare e ottimizzare quanto già implementato nella Rete di prova, in modo che ogni cosa implementata sulla rete del database sia presente anche in versione ridotta.

Per rendere più comprensibile le previsioni di probabilità ottenute, a seguito dell'addestramento della rete e della data in pasto del vettore previsione, ho realizzato un'immagine canvas in cui ogni domanda viene raffigurata con un quadrattino colorato, in base alla previsione risultante (verde se a 1, bianco

a 0, rosso a -1, gradazioni di bianco - verde e bianco - rosso per i valori intermedi).

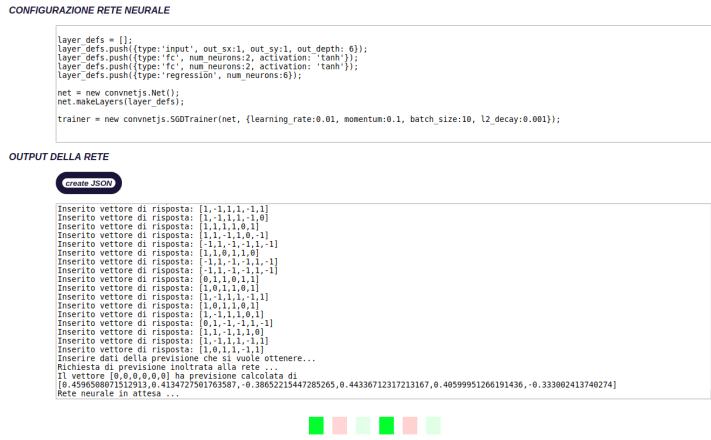


Figura 7: Rete di prova dopo lo sviluppo del canvas per le previsioni.



Figura 8: Rete neurale del database aziendale.

Una prima architettura su cui ho deciso di analizzare i risultati della rete del database a 89 domande di logica, usando quanto ho appreso dalla Rete

di prova e tenendo in conto del numero di vettori di test utilizzati (1245 vettori x 89), è stata la seguente:

```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```

Ho aggiunto un layer e messo un numero di neuroni per layer in modo da formare un romboide.

### 2.2.2 Test e Documentazione

Durante la settimana dal 10/06 al 18/06 ho svolto attività di test e documentazione del lavoro svolto.

#### Test nella Rete di prova

Architettura della rete utilizzata:

```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:6});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:6});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```

- Il vettore [1,1,1,1,1,1] ha previsione calcolata di  
[0.8521066399598267, 0.898137375081856, 0.9993098151218291, 0.792190337086403,  
0.811145866789799, 0.9514731560722426]

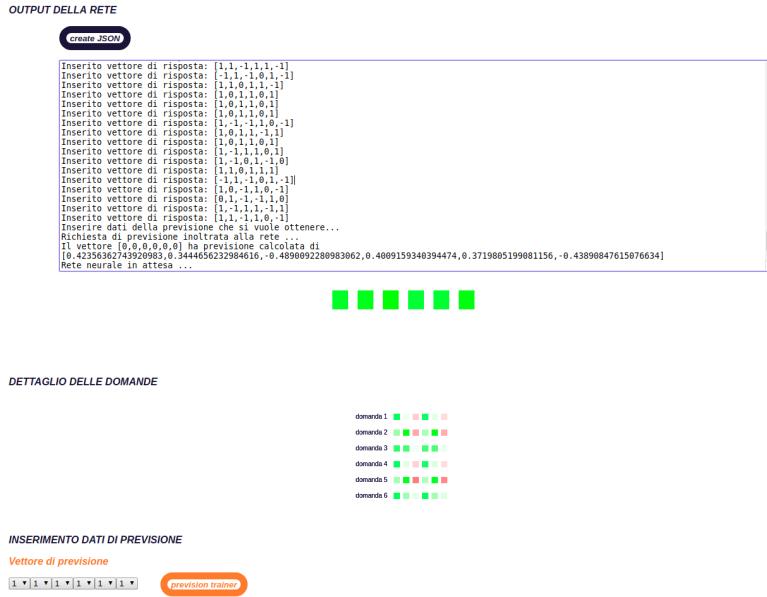


Figura 9: Risultato della rete di prova a seguito di un vettore di previsione [1, 1, 1, 1, 1, 1] in input.

Quanto mostrato dal dettaglio delle domande ha il seguente significato per un candidato:

- se la domanda 1 è settata a 1 (corretta): la rete prevede che la domanda 1 e 4 abbiano una probabilità alta di essere risposte in modo corretto (verde); la 3 e 6 una probabilità non eccessiva di venire risposte in modo sbagliato (rosa attenuato), la 2 e la 5 di non venire nemmeno poste (bianco con qualche minima sfumatura di verde).
- se la domanda 2 è settata a 1 (corretta): la rete prevede che la domanda 1 e 4 abbiano una probabilità non molto alta di essere risposte in modo corretto (bianco con qualche sfumatura di verde); la 3 e 6 una buona probabilità di venire risposte in modo sbagliato (rosa), la 2 e la 5 di venire date in modo corretto (verde).
- se la domanda 3 è settata a 1 (corretta): la rete prevede che la domanda 3 e 6 abbiano una probabilità comunque bassa di essere risposte in modo corretto (bianco con qualche sfumatura di verde); la 1 e 4 con probabilità di venire risposte in modo corretto (verde) perchè più semplici delle domande 3 e 6, la 2 e la 5 di venire risposte correttamente (verde).
- se la domanda 4 è settata a 1 (corretta): la rete prevede un risultato identico a quanto ottenuto dalla domanda 1.

- se la domanda 5 è settata a 1 (corretta): la rete prevede un risultato similare a quanto ottenuto dalla domanda 2. Cambia solo quanto previsto dalle domande 3 e 6 che si presentano con un rosa un pò più intenso, in quanto correlate alla coppia di domande 2 e 5.
- se la domanda 6 è settata a 1 (corretta): la rete prevede un risultato simile a quanto ottenuto dalla domanda 3. La coppia 2 e 5 hanno una probabilità minore di essere date correttamente (bianco con sfumature di verde); ma perchè non correlate alle domande 3 e 6.
- Il vettore  $[-1, -1, -1, -1, -1, -1]$  ha previsione calcolata di  $[0.3440856175367477, -0.5026946644729329, -1.284368009920025, 0.35883842020377565, -0.37844446052773495, -1.1717763012412878]$

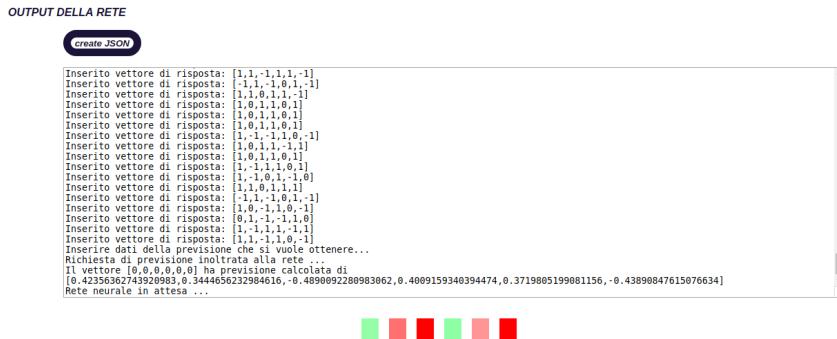


Figura 10: Risultato della rete di prova a seguito di un vettore di previsione  $[-1, -1, -1, -1, -1, -1]$  in input.

- se la domanda 1 è settata a -1 (non corretta): la rete prevede che la domanda 1 e 4 non abbiano una probabilità alta di essere risposte in modo corretto (bianco con sfumature di verde); la 3 e 6 una probabilità molto alta di venire risposte in modo sbagliato (rosso), la 2 e la 5 di non venire nemmeno poste (verde con qualche sfumatura di bianco).

- *se la domanda 2 è settata a -1 (non corretta)*: la rete prevede che la domanda 1 e 4 abbiano una probabilità non molto alta di essere risposte in modo non corretto (verde con qualche sfumatura di bianco); la 3 e 6 una buona probabilità di venire risposte in modo sbagliato (rosa), la 2 e la 5 di venire date in modo non corretto (rosa molto attenuato).
- *se la domanda 3 è settata a -1 (non corretta)*: la rete prevede che la domanda 3 e 6 abbiano una probabilità alta di essere risposte in modo non corretto (rosso); la 1 e 4 con bassa probabilità di venire risposte in modo corretto (bianco con qualche sfumatura di verde) perchè più semplici delle domande 3 e 6, la 2 e la 5 di non venire nemmeno poste o comunque bassa probabilità di venire risposte correttamente (bianco con sfumature di verde).
- *se la domanda 4 è settata a -1 (non corretta)*: la rete prevede un risultato identico a quanto ottenuto dalla domanda 1.
- *se la domanda 5 è settata a -1 (non corretta)*: la rete prevede un risultato similare a quanto ottenuto dalla domanda 2. Cambia solo quanto previsto dalle domande 3 e 6 che si presentano con un rosa un pò meno intenso, in quanto non correlate alla coppia di domande 2 e 5.
- *se la domanda 6 è settata a -1 (non corretta)*: la rete prevede un risultato simile a quanto ottenuto dalla domanda 3. La coppia 2 e 5 hanno una probabilità maggiore di essere date correttamente (bianco con sfumature di verde) perchè non correlate alle domande 3 e 6.

## Test nella Rete del database

### Architetture testate

Durante tutto il periodo ho effettuato una serie di test su molteplici architettura della rete, con gradi di correlazione tra le domande pari al 100% fino ad un differenziale massimo di 5 punti colore, rispetto al canvas risultante per ogni domanda <sup>1</sup>.

Tuttavia non sono riuscita ad individuare un'architettura sufficientemente stabile per prevedere risultati attendibili.

Fattori dominanti, che impediscono la scelta di una configurazione nel caso trattato, sono:

- La molteplicità di dati che hanno aumentato esponenzialmente la complessità di analisi, rispetto alla Rete di prova;

---

<sup>1</sup>valutando i cluster generati

- la creazione stessa del set di input: composto da una probabilità discreta, ma non conosciuta, che il candidato abbia saputo rispondere solo per fortuna.

I fattori sopra citati vengono trattati approfonditamente nella sezione 4.

Architettura della rete utilizzata:

```

● layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```



Figura 11: Risultato della rete del database a seguito di un vettore di previsione  $[1, 1, 1, 1, 1, 1]$  in input.

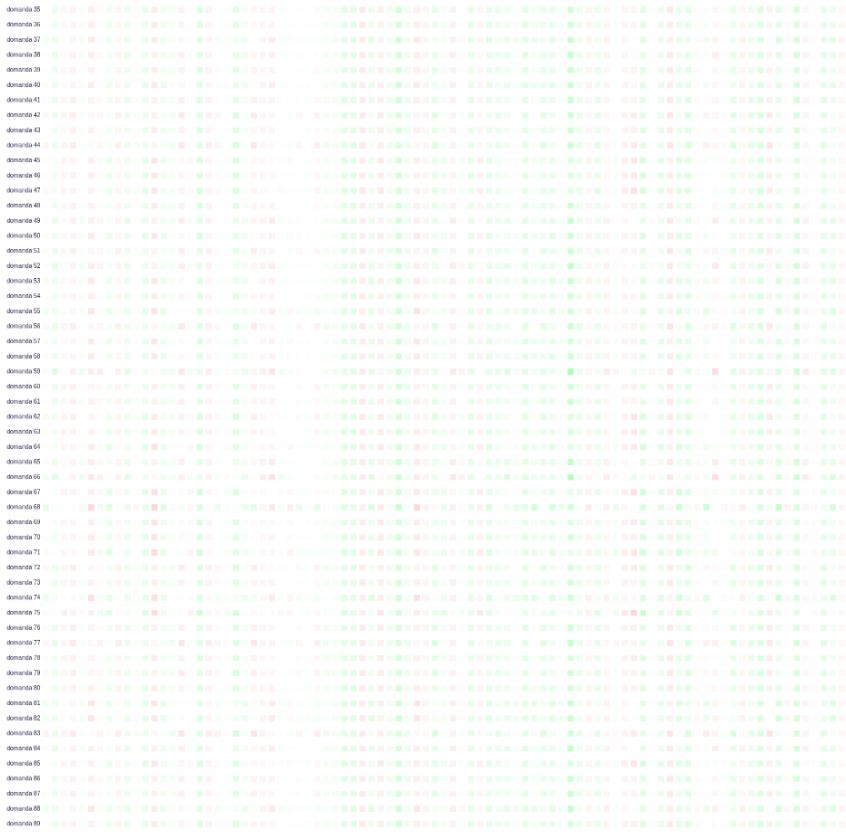


Figura 12: Risultato della rete del database a seguito di un vettore di previsione  $[1, 1, 1, 1, 1, 1]$  in input.

Dagli screen della rete riportati sopra appare come "sembrano" domande:

1. Analisi verticale:

- *semplici* la 18, 22, 34, 35, 37, 39, 41, 44, 48, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 67, 69, 71, 72, 77, 79, 80, 82, 84, 87 e 88. Inoltre di queste sembrano in relazione ancora più stretta le domande 18, 22, 40 e 59.
- *difficili* la 3, 4, 6, 13, 16, 19, 24, 25, 26, 36, 38, 42, 43, 46, 49, 61, 63, 65, 66, 70, 78, 81, 85 e 89. Inoltre di queste sembrano in relazione ancora più stretta le domande 6, 13, 19, 36, 38, 42, 46 e 70.

2. Analisi orizzontale, appaiono accoppiate le seguenti domande:

- 2, 3, 4, 5;
- 7, 8, 9;
- 14, 16;

- 20, 21;
  - 26, 32;
  - 29, 32
  - 33, 34, 35, 36, 38;
  - 39, 41, 43;
  - 46, 48;
  - 49, 52;
  - 50, 53;
  - 72, 79;
  - 81, 82;
  - 86, 87, 88.



Figura 13: Risultato della rete del database a seguito di un vettore di previsione  $[-1, -1, -1, -1, -1, -1]$  in input.

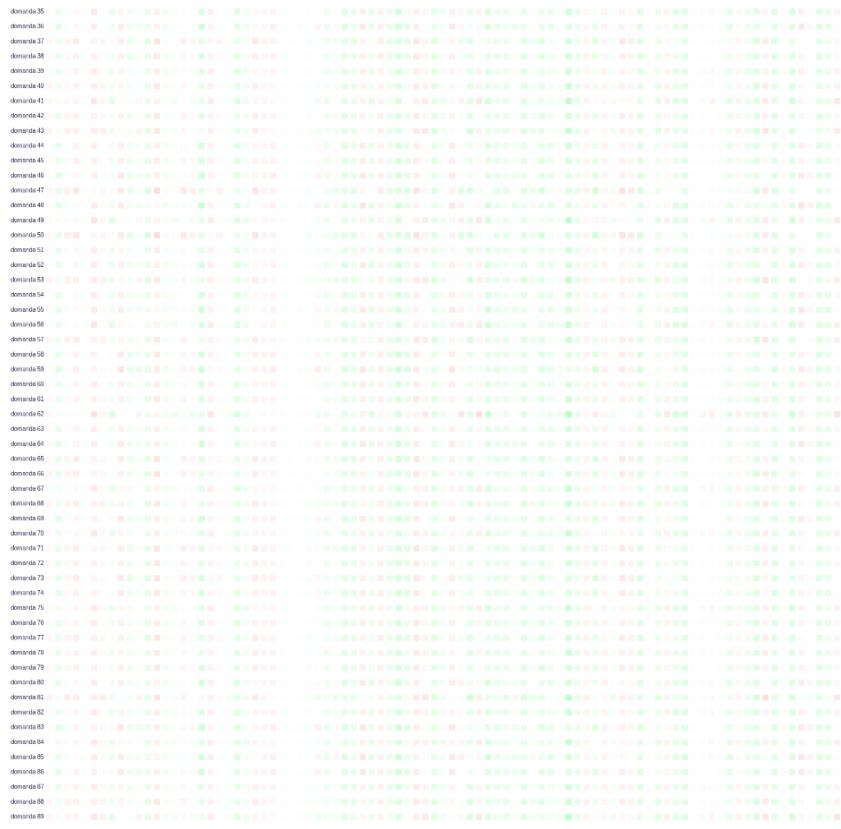


Figura 14: Risultato della rete del database a seguito di un vettore di previsione  $[-1, -1, -1, -1, -1, -1]$  in input.

Dagli screen della rete riportati sopra appare come "sembrano" domande:

1. Analisi verticale:

- *semplici* la 18, 22, 34, 35, 37, 38, 39, 41, 44, 48, 50, 51, 52, 54, 55, 56, 67, 69, 71, 72, 77, 79, 80, 82, 84, 87 e 88. Inoltre di queste sembrano in relazione ancora più stretta le domande 18, 22, 38, 52 e 57.
- *difficili* la 3, 4, 6, 13, 19, 27, 36, 38, 42, 43, 46, 61, 62, 65, 70. Inoltre di queste sembrano in relazione ancora più stretta le domande 6, 13, 19, 36, 38, 42, 46, 61 e 62.

2. Analisi orizzontale appaiono accoppiate le seguenti domande:

- rimaste consistenti con il vettore  $[1, 1, 1, 1, 1, 1]$ .

```
• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
```

```
layer_defs.push({type:'fc', num_neurons:12, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01, momentum:0.1
, batch_size:10, l2_decay:0.001});
```

Noto che aumentando il numero di neuroni sull'unico layer esistente, il valore della domanda corrispondente al vettore della previsione sembra sempre più marcato, segno che la rete "impara troppo" e ricade nel restituire l'immagine stessa del vettore previsione.



Figura 15: Risultato della rete del database a seguito di un vettore di previsione  $[1, 1, 1, 1, 1, 1]$  in input.

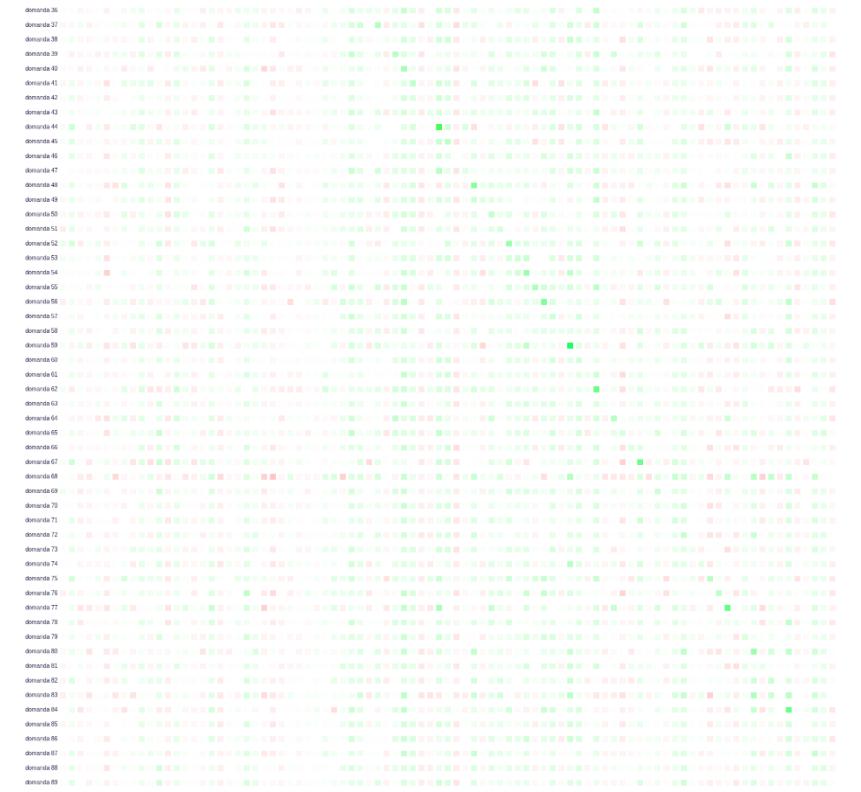


Figura 16: Risultato della rete del database a seguito di un vettore di previsione  $[1, 1, 1, 1, 1, 1]$  in input.

Dagli screen della rete riportati sopra appare come la situazione meno lineare rispetto al caso analizzato precedentemente. Le domande non vengono separate per linee rette; ma per aree di relazione. Appare come "sembrano" domande:

1. Analisi verticale:
  - *semplici* la 18, 22, 34, 35, 37, 39, 41, 44, 48, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 67, 69, 71, 72, 77, 79, 80, 82, 84, 87 e 88. Inoltre di queste sembrano in relazione ancora più stretta le domande 18, 22, 40 e 59.
  - *difficili* la 3, 4, 6, 13, 16, 19, 24, 25, 26, 36, 38, 42, 43, 46, 49, 61, 63, 65, 66, 70, 78, 81, 85 e 89. Inoltre di queste sembrano in relazione ancora più stretta le domande 6, 13, 19, 36, 38, 42, 46 e 70.
2. Analisi orizzontale, appaiono in relazione stretta le seguenti domande:

- Vengono meno le relazioni individuate nella configurazione precedentemente.

Altre architetture sottoposte a test:

- ```
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:12, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```
- ```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:12, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:8, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```
- ```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001})
```

- ```

layer_defs = [];
layer_defs.push
({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```
- ```

layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```
- ```

layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```
- ```

layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});

```

```

layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];

```

```

layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:5, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

```

```

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```

- layer\_defs = [];  
layer\_defs.push({type:'input', out\_sx:1, out\_sy:1, out\_depth:89});  
layer\_defs.push({type:'fc', num\_neurons:18, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:18, activation: 'tanh'});  
layer\_defs.push({type:'regression', num\_neurons:89});  
  
net = new convnetjs.Net();  
net.makeLayers(layer\_defs);  
  
trainer = new convnetjs.SGDTrainer(net, {learning\_rate:0.01,  
momentum:0.1, batch\_size:10, l2\_decay:0.001});
- layer\_defs = [];  
layer\_defs.push({type:'input', out\_sx:1, out\_sy:1, out\_depth:89});  
layer\_defs.push({type:'fc', num\_neurons:18, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:18, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:18, activation: 'tanh'});  
layer\_defs.push({type:'regression', num\_neurons:89});  
  
net = new convnetjs.Net();  
net.makeLayers(layer\_defs);  
  
trainer = new convnetjs.SGDTrainer(net, {learning\_rate:0.01,  
momentum:0.1, batch\_size:10, l2\_decay:0.001});
- layer\_defs = [];  
layer\_defs.push({type:'input', out\_sx:1, out\_sy:1, out\_depth:89});  
layer\_defs.push({type:'fc', num\_neurons:2, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:3, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:2, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:1, activation: 'tanh'});  
layer\_defs.push({type:'regression', num\_neurons:89});  
  
net = new convnetjs.Net();  
net.makeLayers(layer\_defs);  
  
trainer = new convnetjs.SGDTrainer(net, {learning\_rate:0.01,  
momentum:0.1, batch\_size:10, l2\_decay:0.001});
- layer\_defs = [];  
layer\_defs.push({type:'input', out\_sx:1, out\_sy:1, out\_depth:89});  
layer\_defs.push({type:'fc', num\_neurons:2, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:3, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:3, activation: 'tanh'});  
layer\_defs.push({type:'fc', num\_neurons:1, activation: 'tanh'});

```

layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:3, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:6, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:1, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons:8, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:8, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

```

```

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```

Avvalendomi dello studio approfondito del testo delle 89 domande di logica e dai test sopra effettuati, ho tratto le seguenti considerazioni sull'architettura di una rete:

- Un layer non è sufficiente per trattare il problema in esame. Si si aspetta che i dati si aggreghino per argomento: due macro argomenti principali (serie e eulero-venn) e la parte restante, domande senza un vero e proprio tema, concorrente a creare il pavimento della Rete;
- Un numero di neuroni superiore a 11 per hidden layer genera situazioni riconducibili all'overfitting.

L'obiettivo di individuare una configurazione di rete abbastanza "forte", da permettere un riscontro di correlazioni e cluster esistenti, nella stessa, è tuttavia fallito. I risultati ottenuti non sono sufficienti per riuscire a trarre delle conclusioni ragionevoli. Per proseguire il mio studio è stato indispensabile avvalersi di altri strumenti di supporto, come la tecnica di **Principal Component Analysis**.

### **2.3 Funzionalità offerte dall'interfaccia della Rete neurale costruita**

La Rete neurale sviluppata è stata pensata per una persona che deve svolgere analisi su una grande quantità dati, mirata alla previsione dei risultati. L'applicativo realizzato offre le seguenti funzionalità:

- Caricamento dei dati con l'utilizzo di formati *CSV* o *TXT*;
- Possibilità di visualizzare i dati caricati direttamente su pagina web. Ogni risposta viene presentata con codice di test e codice della domanda;
- Possibilità di visualizzare i dati caricati nella Rete mediante formato *JSON*. In questo modo la rete post apprendimento viene salvata ed è possibile visualizzare come ogni nodo pesa ogni variabile. Inoltre l'ultima rete salvata può venire caricata in ogni momento e riutilizzata;
- La Rete neurale è fornita di due textarea:
  - La prima che permette la visualizzazione e la modifica dell'architettura in uso, con indicazione della tipologia di trainer utilizzato. Questa sezione permette la modifica delle variabili di configurazione.
  - La seconda che consente la visualizzazione dei risultati ottenuti dalla Rete.

- È possibile settare la previsione che si vuole ottenere, mediante un'area di inserimento con box a radio. Il valore di default imposto è 0.
- Per ogni previsione svolta viene visualizzato non solo il risultato della stessa, numericamente espresso all'interno della seconda textarea, ma viene anche rappresentato visivamente per mezzo di canvas;
- È offerta la funzionalità di inserimento dei parametri di differenziale con filtro sulla colorazione, che permette l'indicazione dei cluster esistenti;
- È possibile visualizzare per ogni elemento, soggetto alla previsione, il dettaglio della previsione stessa. Questa viene presentata visivamente per mezzo di canvas;
- È possibile eliminare lo storico dei dati stampati nella seconda textarea.

### 2.3.1 Diagrammi dei casi d'uso

#### Attori

Gli *attori* esistenti all'interno del sistema che interagiscono con esso sono:

- **Utente osservatore:** utente che agisce direttamente sulla Rete neurale, in modo da effettuare osservazioni ed analisi dei dati di interesse.

#### Casi d'uso

Ogni caso d'uso è composto dalle seguenti parti:

- **Identificativo:** composto da una stringa testuale, che rappresenta la funzionalità che si vuole modellare, seguita da un codice numerico;
- **Descrizione:** contiene una breve descrizione del caso d'uso;
- **Attori:** l'attore è tutto ciò che è esterno al sistema e con il quale interagisce. Un attore si suddivide in due categorie:
  - **Attore primario:** che interagisce direttamente con il sistema;
  - **Attore secondario:** entità esterna che interagisce con il sistema; ma con l'obiettivo di far raggiungere all'attore principale il suo scopo.
- **Precondizione:** descrive lo stato del sistema prima del verificarsi delle azioni espresse nel caso d'uso;
- **Postcondizione:** descrive lo stato del sistema una volta che si sono svolte le azioni stabilite nel caso d'uso;

- **Scenario principale:** rappresenta in modo puntale, per passi mediante elenco numerato, le azioni del caso d'uso;
- **Estensioni** (Opzionali): ognuno di essi viene eseguito solo quando si avvera una determinata condizione. Se la condizione viene verificata l'esecuzione del caso d'uso, a cui e' collegato, viene interrotta;
- **Inclusioni** (Opzionali): quando due casi d'uso sono tra loro collegati. Viene eseguito sempre e solo dopo che il caso d'uso, a cui è collegato, sia stato eseguito completamente;
- **Generalizzazioni** (Opzionali): specializzazioni di un caso d'uso. Ogni specializzazione è mutualmente accessibile.

#### UC0\_g: Scelta della grandezza del set di domande

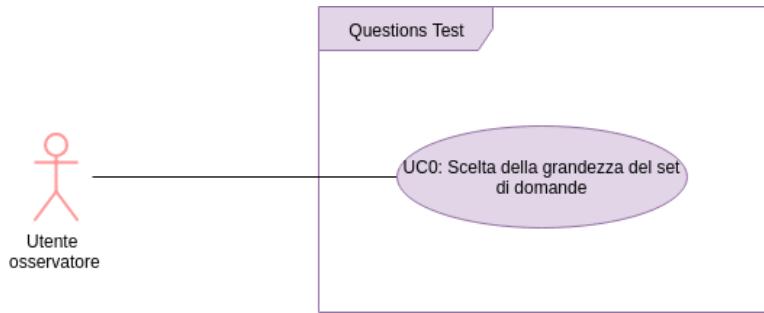


Figura 17: UC0\_g: Operazioni utente - Visione generale della pagina di accesso alla Rete - Questions Test

- **Descrizione:** L'utente sceglie, tra quelle disponibili nell'applicativo, il numero di domande su cui utilizzare la Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente e il sistema non hanno svolto ancora nessuna attività;
- **Postcondizione:** L'utente ha scelto la dimensione del set di domande;
- **Scenario principale:**
  1. L'utente sceglie come dimensione del set di domande 89;
  2. L'utente sceglie come dimensione del set di domande 120;
  3. L'utente conferma la scelta della dimensione.

### UC1-g: Operazioni utente - Visione generale interfacciamento con la Rete neurale

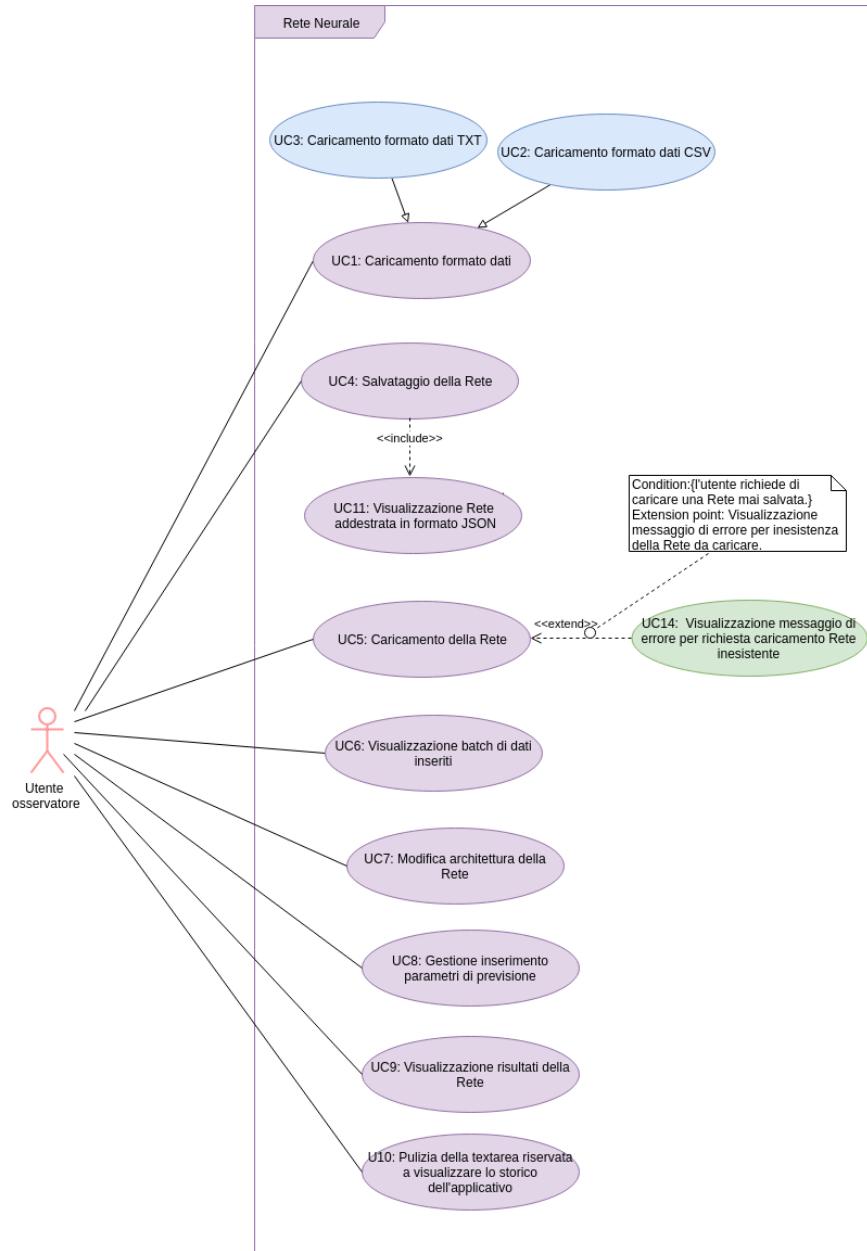


Figura 18: UC1-g: Operazioni utente - Visione generale interfacciamento con la Rete neurale

- **Descrizione:** Il sistema permette all'utente di interfacciarsi con la Rete neurale;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete neurale è in attesa che l'utente effettui almeno un'operazione;
- **Postcondizione:** L'utente si è interfacciato con le funzionalità offerte della Rete neurale;
- **Scenario principale:**
  1. L'utente può effettuare il caricamento dei dati di analisi nella Rete in formato txt o csv (UC1);
  2. L'utente può effettuare il salvataggio di una Rete addestrata (UC4);
  3. L'utente può caricare nel sistema l'ultima Rete salvata (UC5);
  4. L'utente può visualizzare il batch di dati immessi nella Rete (UC6);
  5. L'utente può modificare la configurazione dell'architettura di Rete (UC7);
  6. L'utente può effettuare l'inserimento dei parametri di previsione (UC8);
  7. L'utente può visualizzare i risultati della Rete, ottenuti sui dati caricati e addestrati, a seguito delle operazioni di previsione (UC9);
  8. L'utente può procedere all'eliminazione dello storico della Rete contenuto all'interno della textarea dedicata (UC10).

#### **UC1: Caricamento formato dati**

- **Descrizione:** Il sistema permette all'utente di poter caricare i dati su cui effettuare l'analisi usando formato csv o txt;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete neurale è in attesa che l'utente effettui l'operazione di caricamento del file di dati;
- **Postcondizione:** L'utente ha potuto caricare il file di dati di suo interesse;
- **Scenario principale:**
  1. L'utente effettua il caricamento di un file dati;

2. L'utente conferma l'operazione di caricamento dati (UC1.1).

#### **Generalizzazioni:**

1. L'utente può effettuare il caricamento di un formato di dati csv (UC2);
2. L'utente può effettuare il caricamento di un formato di dati txt (UC3).

#### **UC2: Caricamento formato dati CSV**

- **Descrizione:** Il sistema permette all'utente di poter caricare i dati su cui effettuare l'analisi usando formato csv;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete neurale è in attesa che l'utente effettui l'operazione di caricamento del file di dati;
- **Postcondizione:** L'utente ha potuto caricare il file di dati di suo interesse nel formato csv;
- **Scenario principale:** L'utente sceglie il file di dati che vuole caricare nella Rete.

#### **UC3: Caricamento formato dati TXT**

- **Descrizione:** Il sistema permette all'utente di poter caricare i dati su cui effettuare l'analisi usando formato txt;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete neurale è in attesa che l'utente effettui l'operazione di caricamento del file di dati;
- **Postcondizione:** L'utente ha potuto caricare il file di dati di suo interesse nel formato txt;
- **Scenario principale:** L'utente sceglie il file di dati che vuole caricare nella Rete.

#### **UC1.1: Conferma operazione di caricamento**

- **Descrizione:** L'utente conferma il file dati scelto con cui addestrare la Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha scelto il file dati da caricare;

- **Postcondizione:** L'utente ha potuto caricare il file di dati di suo interesse;
- **Scenario principale:**
  1. L'utente ha confermato il file dati scelto con l'apposto pulsante.
- **Estensioni:**
  1. Il browser in uso dall'utente non supporta lo standard HTML5. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC1.1.1);
  2. L'utente ha effettuato il caricamento di un formato di dati non conforme al formato csv o txt. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC1.1.2).
  3. L'utente ha effettuato il caricamento di un formato di dati che non rispetta la configurazione di Rete scelta. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC1.1.3).

#### **UC1.1.1: Visualizzazione messaggio di errore di mancata idoneità del browser in uso**

- **Descrizione:** Il file nel formato corretto viene caricato nel sistema, che non è in grado di processare alcuna informazione in quanto il browser, in uso dall'utente, non supporta lo standard di markup HTML5;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente non ha ancora caricato il file nella Rete;
- **Postcondizione:** Il sistema produce un messaggio d'errore per l'utente; con il quale quest'ultimo viene informato nella non conformità del browser in uso, con gli standard minimi richiesti, per portare a termine l'operazione di caricamento del file nella Rete con successo;
- **Scenario principale:** Il file scelto dall'utente viene caricato all'interno della Rete.

#### **UC1.1.2: Visualizzazione messaggio di errore per formato di dati non coerente con le aspettative**

- **Descrizione:** Il file che viene caricato nella rete non è del formato richiesto corretto;
- **Attori:** Utente osservatore;

- **Precondizione:** L'utente non ha ancora caricato il file nella Rete;
- **Postcondizione:** Il sistema produce un messaggio d'errore per l'utente; con il quale quest'ultimo viene informato della non conformità del formato di dati scelto;
- **Scenario principale:** Il file viene caricato all'interno della Rete.

**UC1.1.3: Visualizzazione messaggio di errore di inconsistenza del set di dati rispetto alla configurazione della Rete**

- **Descrizione:** Il file nel formato corretto viene caricato nel sistema, che non è in grado di processare alcuna informazione in quanto, il numero di domande del set di dati è differente alla configurazione della Rete, imposta inizialmente dall'utente
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha richiesto alla Rete di caricare un file;
- **Postcondizione:** Il sistema produce un messaggio d'errore per l'utente; con il quale quest'ultimo viene informato della non conformità del file scelto, rispetto al numero di domande da processare in rapporto alla configurazione della Rete;
- **Scenario principale:** Il file scelto dall'utente viene caricato all'interno della Rete.

**UC4: Salvataggio della Rete**

- **Descrizione:** L'utente richiede che i dati caricati all'interno della Rete vengano salvati nello stato in cui si trovano;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha caricato un file dati nella Rete e quest'ultima ne ha provveduto all'addestramento;
- **Postcondizione:** L'utente ha gestito i dati caricati all'interno della Rete in modo da salvare la Rete;
- **Scenario principale:** L'utente effettua l'operazione di salvataggio della Rete mediante l'apposito pulsante.
- **Inclusioni:**
  1. La Rete salvata viene visualizzata sulla textarea riservata, presentando i dati della rete addestrata in formato JSON (UC11).

### **UC11: Visualizzazione Rete addestrata in formato JSON**

- **Descrizione:** L'utente visualizza in formato JSON i dati della Rete sottoposti ad addestramento;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha effettuato il salvataggio della Rete;
- **Postcondizione:** L'utente ha visualizzato la Rete salvata, tradotta in formato JSON, all'interno della textarea riservata dell'applicativo.
- **Scenario principale:** L'utente visualizza i dati addestrati in formato JSON.

### **UC5: Caricamento della Rete**

- **Descrizione:** L'utente richiede che la Rete salvata precedentemente venga ricaricata e preparata all'uso nella Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha salvato una Rete precedentemente;
- **Postcondizione:** L'utente ha gestito i dati caricati all'interno della Rete in modo da ricaricare una Rete precedentemente salvata;
- **Scenario principale:** L'utente effettua il caricamento della Rete precedente salvata mediante l'apposito pulsante.
- **Estensioni:**
  1. L'utente ha richiesto il caricamento di una Rete mai precedentemente salvata. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC14).

### **UC14: Visualizzazione messaggio di errore per richiesta caricamento Rete inesistente**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui abbia richiesto il caricamento di una Rete quando non ne è ancora stata salvata nessuna;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha richiesto il caricamento dei dati della Rete precedente;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'inesistenza di una Rete da caricare;
- **Scenario principale:** All'utente viene indicato l'errore.

## **UC6: Visualizzazione batch di dati inseriti**

- **Descrizione:** L'utente visualizza i dati caricati nella Rete all'interno di una textarea riservata, all'interno della pagina web dell'applicativo. Ne può ottenere un maggior dettaglio visualizzando gli stessi su una pagina web riservata, in cui vengono evidenziati i significati dei dati mostrati;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha effettuato il caricamento dei dati nella Rete in uno dei formati idonei;
- **Postcondizione:** L'utente visualizza il batch di dati caricati nella Rete;
- **Scenario principale:**
  1. L'utente visualizza i dati caricati nella Rete all'interno della pagina web dell'applicativo stesso (UC6.1);
  2. L'utente visualizza i dati caricati nella Rete su una pagina web a parte (UC6.2).

### **UC6.1: Visualizzazione dei dati caricati all'interno della pagina web riservata all'applicativo**

- **Descrizione:** L'utente visualizza i dati caricati nella Rete all'interno di una textarea riservata, all'interno della pagina web dell'applicativo, che ha il compito di tenere traccia di tutte le operazioni svolte nella Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha effettuato il caricamento dei dati nella Rete, in uno dei formati idonei;
- **Postcondizione:** L'utente ha visualizzato all'interno della textarea i vettori di dati caricati nella Rete;
- **Scenario principale:** L'utente visualizza i dati caricati nella Rete all'interno della textarea dell'applicativo prima che il sistema proceda con l'addestramento.

### **UC6.2: Visualizzazione dei dati caricati all'interno della Rete su pagina web a parte**

- **Descrizione:** L'utente visualizza i dati caricati nella Rete su una pagina web riservata ad adempire tale mansione. I dati caricati vengono presentati con indicazione del codice di test e di domanda a cui ognuno fa capo. La funzionalità ha lo scopo di rendere più agevole lo svolgimento di osservazioni sui dati;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha effettuato il caricamento dei dati nella Rete in uno dei formati idonei;
- **Postcondizione:** L'utente ha visualizzato all'interno di una pagina web riservata i vettori di dati caricati nella Rete;
- **Scenario principale:** L'utente visualizza i dati caricati nella Rete all'interno di una pagina web reindirizzata dall'applicativo.

#### **UC7: Modifica architettura della Rete**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale modificando i layer presenti e il loro contenuto.
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;
- **Scenario principale:**
  1. L'utente effettua sulla Rete operazioni di modifica dell'architettura concesse (UC7.1);
  2. L'utente effettua sulla Rete operazioni di modifica dell'architettura non concesse (U7.2);
  3. L'utente conferma l'operazione di modifica dell'architettura di Rete (UC7.3).

#### **UC7.1: Gestione operazioni di configurazione dell'architettura concesse**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale. Esso può procedere a modificare il numero di neuroni per layer, impostare dei nuovi layers, eliminare i layers intermedi esistenti, modificare/eliminare la funzione attivazione e ridefinire i parametri del trainer della Rete;

- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;
- **Scenario principale:**
  1. L'utente ha aggiunto un nuovo layer intermedio nella Rete (UC7.1.1);
  2. L'utente ha eliminato un layer intermedio esistente nella Rete (UC7.1.2);
  3. L'utente ha modificato un layer intermedio (UC7.1.3);
  4. L'utente ha modificato i parametri del trainer della Rete (UC7.1.4).

#### **UC7.1.1: Aggiunta layer intermedio**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale. Esso può procedere ad aggiungere un layer intermedio nella Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità aggiungendo un nuovo layer intermedio;
- **Scenario principale:**
  1. L'utente ha aggiunto un layer intermedio alla Rete:
    - (a) L'utente effettua l'inserimento del parametro di tipo;
    - (b) L'utente effettua l'inserimento del numero di neuroni;
    - (c) L'utente può indicare o meno la funzione attivazione;

#### **UC7.1.2: Eliminazione layer intermedio esistente**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale. Esso può procedere ad eliminare un layer intermedio presente nella Rete;
- **Attori:** Utente osservatore;

- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità eliminando un layer intermedio;
- **Scenario principale:**

1. L'utente ha eliminato un layer intermedio esistente nella Rete:
  - L'utente elimina il parametro di tipo;
  - L'utente elimina il numero di neuroni;
  - L'utente elimina, se presente, la funzione attivazione.

#### **UC7.1.3: Modifica layer intermedio esistente**

- **Descrizione:** L'utente ha la possibilità di modificare un layer intermedio presente nella Rete
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità provvedendo a ridefinire il contenuto di un layer intermedio;
- **Scenario principale:**

1. L'utente ha ridefinito un layer intermedio presente nella Rete:
  - L'utente può modificare il numero di neuroni presenti nella Rete;
  - L'utente può modificare/aggiungere/eliminare la funzione attivazione.

#### **UC7.1.4: Modifica di trainer**

- **Descrizione:** L'utente ha la possibilità di configurare la dichiarazione del trainer della Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente non ha ancora effettuato alcuna operazione e la Rete desiderata ancora non esiste;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità definendo il trainer;

- **Scenario principale:**

1. L'utente ha riconfigurato il trainer della Rete:
  - (a) L'utente può modificare il parametro di learning\_rate;
  - (b) L'utente può modificare il parametro momentum;
  - (c) L'utente può modificare il parametro batch\_size;
  - (d) L'utente può modificare il parametro l2\_decay.

**UC7.2: Gestione operazioni di configurazione dell'architettura non concesse**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale in base alle proprie esigenze e ai valori delle variabili di cui necessità per compiere le osservazioni sulla Ret; ma non tutte le operazioni di modifica/cancellazione possono venire compiute.

- **Attori:** Utente osservatore;

- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;

- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;

- **Scenario principale:**

1. Le operazioni non concesse all'utente riguardano le seguenti operazioni:
  - (a) Modifica del layer di input (UC7.2.1);
  - (b) Cancellazione del layer di input (UC7.2.2);
  - (c) Modifica del layer di output (UC7.2.3);
  - (d) Cancellazione del layer di output (UC7.2.4);
  - (e) Cancellazione parziale di un layer intermedio (UC7.2.5).

**UC7.2.1: Modifica layer di input**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale in base alle proprie esigenze, ed è interessato ad effettuando la modifica del layer di input;

- **Attori:** Utente osservatore;

- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;

- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;
- **Scenario principale:**

1. Le operazioni non concesse all'utente riguardano la modifica nel layer di input:
  - (a) Modifica del tipo di input;
  - (b) Modifica della profondità della Rete;
  - (c) Modifica della larghezza della Rete;
  - (d) Modifica dell'altezza della Rete.

#### **UC7.2.2: Cancellazione del layer di input**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale in base alle proprie esigenze, ed è interessato ad effettuando la cancellazione del layer di input;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;
- **Scenario principale:**

1. Le operazioni non concesse all'utente riguardano la cancellazione del layer di input:
  - (a) Cancellazione del tipo di input;
  - (b) Cancellazione della profondità della Rete;
  - (c) Cancellazione della larghezza della Rete;
  - (d) Cancellazione dell'altezza della Rete.

#### **UC7.2.3: Modifica layer di output**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale in base alle proprie esigenze, ed è interessato ad effettuando la modifica del layer di output;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione sulla essa;

- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;
- **Scenario principale:**
  1. Le operazioni non concesse all'utente riguardano la modifica nel layer di output:
    - (a) Modifica del tipo di regressione;
    - (b) Modifica del numero di neuroni.

#### **UC7.2.4: Cancellazione del layer di output**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale in base alle proprie esigenze, ed è interessato ad effettuando la cancellazione del layer di output;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;
- **Scenario principale:**
  1. Le operazioni non concesse all'utente riguardano la cancellazione del layer di output:
    - (a) Cancellazione del tipo di regressione;
    - (b) Cancellazione del numero di neuroni.

#### **UC7.2.5: Cancellazione parziale di un layer intermedio esistente**

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale in base alle proprie esigenze, ed è interessato ad effettuando la cancellazione di alcuni campi di un layer intermedio esistente;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete desiderata dall'utente non esiste, e di conseguenza l'utente non ha ancora effettuato alcuna operazione su di essa;
- **Postcondizione:** L'utente ha provveduto a configurare la Rete in base alle proprie necessità;

- **Scenario principale:**

1. Le operazioni non concesse all’utente riguardano la cancellazione del layer intermedio:
  - (a) Cancellazione del tipo del layer intermedio;
  - (b) Cancellazione del numero di neuroni;
  - (c) Cancellazione della funzione attivazione, se presente.

### **UC7.3: Conferma dell’operazione di modifica della Rete**

- **Descrizione:** L’utente conferma la configurazione della Rete presentata nella textarea dedicata;

- **Attori:** Utente osservatore;

- **Precondizione:** La Rete desiderata dall’utente non esiste, e di conseguenza l’utente non ha ancora effettuato alcuna operazione sul di essa;

- **Postcondizione:** L’utente ha provveduto a configurare la Rete in base alle proprie necessità;

- **Scenario principale:**

1. L’utente ha confermato la configurazione di Rete.

- **Estensioni:**

1. L’utente ha tentato di modificare il type input sul layer di input della Rete. Viene ritornato un messaggio esplicativo che indica l’impossibilità di terminare l’operazione (UC7.4);
2. L’utente ha tentato di modificare la larghezza dichiarata (out\_sx) sul layer di input della Rete. Viene ritornato un messaggio esplicativo che indica l’impossibilità di terminare l’operazione (UC7.5);
3. L’utente ha tentato di modificare l’altezza dichiarata (out\_sy) sul layer di input della Rete. Viene ritornato un messaggio esplicativo che indica l’impossibilità di terminare l’operazione (UC7.6);
4. L’utente ha tentato di modificare la profondità dichiarata (out\_depth) sul layer di input della Rete. Viene ritornato un messaggio esplicativo che indica l’impossibilità di terminare l’operazione (UC7.7);
5. L’utente ha tentato di modificare il tipo di connessione fulled connected su un layer intermedio della Rete. Viene ritornato un messaggio esplicativo che indica l’impossibilità di terminare l’operazione (UC7.8);

6. L'utente ha tentato di cancellare il numero di neuroni presenti su un layer intermedio della Rete. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC7.9);
7. L'utente ha tentato di inserire un numero di neuroni presenti su un layer intermedio della Rete minore o uguale a 0. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC7.10);
8. L'utente ha tentato di modificare la funzione di attivazione su un layer intermedio della Rete, con una dichiarazione di funzione non valida. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC7.11);
9. L'utente ha tentato di modificare il tipo regressione presente sul layer di output della Rete. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC7.12);
10. L'utente ha tentato di modificare il numero di neuroni presente sul layer di output della Rete. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC7.13).

**UC7.6: Visualizzazione messaggio di errore di type input sul layer di input**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui abbia modificato nel layer di input il tipo dello stesso;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha fornito un tipo nel layer di input
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo richiesto con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.7: Visualizzazione messaggio di errore per la larghezza dichiarata (out\_sx) sul layer di input**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui abbia modificato nel layer di input la larghezza dichiarata;
- **Attori:** Utente osservatore;

- **Precondizione:** L'utente ha fornito la dimensione della larghezza della componente base della Rete Vol;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità della dimensione richiesta con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.8: Visualizzazione messaggio di errore per altezza dichiarata (out\_sy) sul layer di input**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui abbia modificato nel layer di input l'altezza dichiarata;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha fornito la dimensione dell'altezza della componente base della Rete Vol;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità della dimensione richiesta con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.9: Visualizzazione messaggio di errore per la profondità (out\_depth) dichiarata input sul layer di input**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui abbia modificato nel layer di input la profondità dichiarata;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha fornito la dimensione della profondità della componente base della Rete Vol;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità della dimensione richiesta con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.10: Visualizzazione messaggio di errore per type fulled connected sui layers intermedi**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui abbia inserito in uno dei layers intermedi un tipo diverso da fulled connected;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha fornito il tipo in un layer intermedio;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo richiesto con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.11: Visualizzazione messaggio di errore per cancellazione numero di neuroni sui layers intermedi**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso in cui effettuato la cancellazione in uno dei layers intermedi del numero di neuroni presenti;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha eliminato il numero di neuroni da un layer intermedio;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità dell'operazione richiesta con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.12: Visualizzazione messaggio di errore per inserimento numero di neuroni sui layers intermedi minore o uguale a 0**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso abbia inserito/modificato il numero di neuroni presenti in un layer intermedio con un valore minore o uguale a 0;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha inserito il numero di neuroni in un layer intermedio;

- **Postcondizione:** Il sistema ha fornito all’utente la visualizzazione di un messaggio di errore che indica l’incompatibilità del valore definito con la natura della Rete;
- **Scenario principale:** All’utente viene indicato l’errore con l’indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.13: Visualizzazione messaggio di errore per la funzione attivazione presente sui layers intermedi**

- **Descrizione:** L’utente visualizza un messaggio di errore nel caso abbia inserito/modificato la funzione di attivazione su un layer intermedio con parametro diverso da uno dei desiderati;
- **Attori:** Utente osservatore;
- **Precondizione:** L’utente ha inserito una funzione attivazione in un layer intermedio;
- **Postcondizione:** Il sistema ha fornito all’utente la visualizzazione di un messaggio di errore che indica l’inesistenza della funzione attivazione richiesta per la Rete;
- **Scenario principale:** All’utente viene indicato l’errore con l’indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.14: Visualizzazione messaggio di errore per type regression sul layer di output**

- **Descrizione:** L’utente visualizza un messaggio di errore nel caso abbia modificato il tipo dichiarato nel layer di output
- **Attori:** Utente osservatore;
- **Precondizione:** L’utente ha modificato il tipo presente nel layer di output;
- **Postcondizione:** Il sistema ha fornito all’utente la visualizzazione di un messaggio di errore che indica l’incompatibilità del valore del tipo richiesto con la natura della Rete;
- **Scenario principale:** All’utente viene indicato l’errore con l’indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC7.15: Visualizzazione messaggio di errore per numero di neuroni sul livello di output**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso abbia modificato il numero di neuroni dichiarati nel layer di output
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha modificato il numero di neuroni presenti nel layer di output;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo di neuroni richiesti con la natura della Rete;
- **Scenario principale:** All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

#### **UC8: Gestione inserimento parametri di previsione**

- **Descrizione:** L'utente inserisce i parametri necessari per effettuare previsione sui dati addestrati della Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete è addestrata con i dati caricati dall'utente;
- **Postcondizione:** L'utente ha ottenuto la previsione sui dati addestrati in base ai parametri passati in input alla Rete;
- **Scenario principale:**
  1. L'utente ha inserito i parametri necessari per effettuare una clusterizzazione sui dati predetti dall'applicativo (UC8.1);
  2. L'utente ha selezionato per ogni domanda una delle risposte, tra quelle concesse, in modo da generare un vettore previsione con cui effettuare la previsione sui dati addestrati dalla Rete (UC8.2);
  3. L'utente ha confermato le operazioni dando il via al calcolo dei risultati di previsione (UC8.3).

#### **UC8.1: Inserimento parametri di differenziale di accoppiamento per effettuare clusterizzazione della previsione**

- **Descrizione:** L'utente inserisce i parametri necessari per effettuare clusterizzazione una volta ottenuti dalla Rete i dati di previsione;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete è addestrata con i dati caricati dall'utente;

- **Postcondizione:** L'utente ha inserito i parametri di differenziale di accoppiamento necessari;
- **Scenario principale:**
  1. Vengono inseriti i tre parametri di clusterizzazione:
    - (a) L'utente può inserire il primo parametro differenziale;
    - (b) L'utente può inserire il secondo parametro differenziale;
    - (c) L'utente può inserire il terzo parametro differenziale.

#### **UC8.2: Inserimento dei parametri del vettore previsione**

- **Descrizione:** L'utente inserisce il valore delle domande per effettuare previsione sui dati addestrati della Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete è addestrata con i dati caricati dall'utente;
- **Postcondizione:** L'utente ha inserito le risposte alle domande nel vettore previsione;
- **Scenario principale:**
  1. L'utente sceglie il contenuto del vettore previsione:
    - (a) L'utente sceglie la risposta alla domanda  $n = 1, \dots, \dim$  — dim numero di domande scelte, 89 o 120 con uno dei seguenti valori: 0 risposta non data, -1 risposta sbagliata e 1 risposta corretta.

#### **UC8.3: Conferma operazioni di previsione inserite**

- **Descrizione:** L'utente conferma i dati di differenziali di accoppiamento e il vettore di previsione inserito;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha inserito i parametri su cui effettuare previsione su una Rete' addestrata;
- **Postcondizione:** L'utente ha confermato i dati inseriti e innescato il calcolo della previsione;
- **Scenario principale:** L'utente conferma i differenziali di accoppiamento e il vettore previsione inseriti nell'applicativo. **Estensioni:**
  1. L'utente ha inserito per uno o più dei parametri di clusterizzazione un valore stringa. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC8.4);

2. L'utente ha inserito per uno o più dei parametri di clusterizzazione un valore vuoto. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC8.5);
3. L'utente ha inserito per uno o più dei parametri di clusterizzazione un numero negativo. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC8.6);
4. L'utente ha inserito per uno o più dei parametri di clusterizzazione un numero superiore a 255. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC8.7).

**UC8.4: Visualizzazione messaggio di errore inserimento valore stringa**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso abbia inserito come parametro di clusterizzazione un valore stringa;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha inserito come differenziale di accoppiamento un valore stringa;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo di valore inserito;
- **Scenario principale:** All'utente viene indicato l'esistenza di un'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

**UC8.5: Visualizzazione messaggio di errore inserimento valore vuoto**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso abbia inserito come parametro di clusterizzazione un valore vuoto;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha inserito come differenziale di accoppiamento un valore vuoto;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo di valore inserito;
- **Scenario principale:** All'utente viene indicato l'esistenza di un'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

### **UC8.6: Visualizzazione messaggio di errore inserimento numero negativo**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso abbia inserito come parametro di clusterizzazione un numero negativo;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha inserito come differenziale di accoppiamento un numero negativo;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo di valore inserito;
- **Scenario principale:** All'utente viene indicato l'esistenza di un'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

### **UC8.7: Visualizzazione messaggio di errore inserimento numero positivo massimo**

- **Descrizione:** L'utente visualizza un messaggio di errore nel caso abbia inserito come parametro di clusterizzazione un numero positivo superiore a 255 (massima parametro  $rgb^2$  possibile);
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha inserito come differenziale di accoppiamento un numero positivo superiore a 255;
- **Postcondizione:** Il sistema ha fornito all'utente la visualizzazione di un messaggio di errore che indica l'incompatibilità del tipo di valore inserito;
- **Scenario principale:** All'utente viene indicato l'esistenza di un'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

### **UC9: Visualizzazione risultati della Rete**

- **Descrizione:** L'utente visualizza i risultati ottenuti dalla Rete a seguito della richiesta, andata a buon fine, di previsione sui dati addestrati della Rete;
- **Attori:** Utente osservatore;

---

<sup>2</sup>modello di colori di tipo additivo: come somma dei tre colori Rosso (Red), Verde (Green) e Blu (Blue).

- **Precondizione:** La Rete è addestrata e con richiesta di previsione pendente;
- **Postcondizione:** L'utente ha visualizzato i risultati di previsione ottenuti dalla Rete;
- **Scenario principale:**
  1. Visualizzazione nella textarea riservata dell'applicativo, dei risultati di previsione ottenuti, espressi in termini numerici su ogni domanda (UC9.1);
  2. Visualizzazione nella textarea riservata dell'applicativo, dei risultati di previsione ottenuti, espressi in codice rgb su ogni domanda (UC9.2);
  3. Visualizzazione nella textarea riservata dell'applicativo, dei cluster delle domande generati sulla base dei parametri di differenziale di accoppiamento, passati in input e applicati alle previsioni espresse in codice rgb (UC9.3);
  4. Visualizzazione generale, mediante tecnologia canvas, dei risultati delle previsioni, espressi in termini rgb su ogni domanda (UC9.4).

#### **UC9.1: Visualizzazione risultati di previsione in termini numerici**

- **Descrizione:** L'utente visualizza i risultati di previsione in termini numerici (calcolati sul vettore previsione della Rete addestrata) all'interno della textarea riservata;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete è addestrata e ha svolto la funzione di previsione;
- **Postcondizione:** L'utente ha visualizzato i risultati di previsione numerici, ottenuti dalla Rete, all'interno della textarea riservata;
- **Scenario principale:** Visualizzazione dei valori di previsione, ottenuti dal vettore di previsione passato in input, stampati all'interno della textarea dedicata a contenere lo storico delle osservazione di Rete.

#### **UC9.2: Visualizzazione risultati di previsione in termini rgb**

- **Descrizione:** L'utente visualizza i risultati di previsione convertiti in rgb dai valori numerici (calcolati sul vettore previsione della Rete addestrata) all'interno della textarea riservata;
- **Attori:** Utente osservatore;

- **Precondizione:** La Rete è addestrata e ha svolto la funzione di previsione;
- **Postcondizione:** L'utente ha visualizzato i risultati di previsione in codice rgb, ottenuti dalla Rete, all'interno della textarea riservata;
- **Scenario principale:** Visualizzazione dei valori di previsione, ottenuti dal vettore di previsione passato in input ed espressi in codice rgb, stampati all'interno della textarea dedicata a contenere lo storico delle osservazioni di Rete.

#### **UC9.3: Visualizzazione risultati di clusterizzazione**

- **Descrizione:** L'utente visualizza, all'interno della textarea dedicata, la clusterizzazione dei dati calcolata sulla previsione espressa in codice rgb;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete è addestrata e ha svolto la funzione di previsione;
- **Postcondizione:** L'utente ha visualizzato la clusterizzazione ottenuta dalla Rete, all'interno della textarea riservata;
- **Scenario principale:** Visualizzazione della clusterizzazione delle domande, generata dalla previsione espressa in codice rgb.

#### **UC9.4: Visualizzazione dei risultati di previsione mediante canvas**

- **Descrizione:** L'utente visualizza i risultati ottenuti dalla Rete, mediante canvas applicati alle previsioni espresse in termini rgb, a seguito della richiesta di previsione sui dati addestrati della Rete;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete è addestrata e con richiesta di previsione pendente;
- **Postcondizione:** L'utente ha visualizzato i risultati di previsione ottenuti dalla Rete rappresentati per mezzo di canvas;
- **Scenario principale:**
  1. Visualizzazione di un quadrato per ogni domanda in analisi con gradazione verde, bianco, rosso; sulla base del risultato di previsione ottenuto a seguito dell'addestramento della Rete e del vettore previsione inoltrato in input;

2. Visualizzazione di dettaglio, mediante tecnologia canvas, della previsione ottenuta (UC9.4.1).

**UC9.4.1: Visualizzazione dei risultati di previsione in dettaglio mediante canvas**

- **Descrizione:** L'utente visualizza i risultati ottenuti dalla Rete, mediante canvas applicata alle previsioni espresse in termini rgb, su un vettore previsione che pone come domanda corretta solo la domanda in esame;
- **Attori:** Utente osservatore;
- **Precondizione:** La Rete ha effettuato previsione sui dati;
- **Postcondizione:** L'utente ha visualizzato i risultati di previsione di dettaglio ottenuti dalla Rete rappresentati per mezzo di canvas;
- **Scenario principale:**

1. Visualizzazione di un quadrato, tanti quante sono le domande per ogni domanda in analisi, con gradazione verde, bianco, rosso. La gradazione è calcolata sul risultato di previsione ottenuto a seguito dell'addestramento della Rete e del vettore previsione che pone corretta la domanda in esame e sbagliate tutte le rimanenti.

**UC10: Pulizia della textarea riservata a visualizzare lo storico dell'applicativo**

- **Descrizione:** L'utente effettua la cancellazione dello storico in visualizzazione nella textarea riservata;
- **Attori:** Utente osservatore;
- **Precondizione:** L'utente ha effettuato nella Rete alcune o anche nessuna operazione;
- **Postcondizione:** L'utente visualizza la textarea pulita dallo storico delle operazioni effettuate dalla Rete;
- **Scenario principale:** L'utente visualizza l'area di lavoro priva di informazioni sui dati elaborati fino a quel momento.

### 2.3.2 Requisiti

#### Classificazione dei requisiti

I requisiti sono classificati come segue:

**R[Importanza][Tipologia][Codice]**

- **Importanza:** ogni requisito può appartenere solo ad una delle classi di importanza elencate di seguito:
  - **O (Requisito Obbligatorio):** requisito fondamentale per la corretta realizzazione del progetto;
  - **D (Requisito Desiderabile):** requisito non fondamentale al progetto, ma il cui soddisfacimento comporterebbe una maggiore completezza del prodotto;
  - **F (Requisito Facoltativo):** requisito non richiesto per il corretto funzionamento del prodotto, ma che se incluso porterebbe valore aggiunto al progetto. Prima di soddisfare tale categoria di requisiti è indispensabile svolgere analisi di tempi e costi, per evitare ritardi nella consegna e/o costi superiori a quelli preventivati.
- **Tipologia:** Di seguito sono riportate le tipologie di requisito:
  - **V:** identifica un requisito di Vincolo, ovvero descrive vincoli sui servizi offerti dal sistema;
  - **F:** identifica un requisito Funzionale, ovvero descrive servizi o funzioni offerti dal sistema;
  - **Q:** identifica un requisito di Qualità, ovvero descrive i vincoli di qualità da realizzare quali manutenibilità, sicurezza, portabilità, disponibilità, ecc...,
- **Codice:** Ogni requisito è formato da un codice numerico progressivo che lo identifica in modo univoco.

**[IDBase](.[IDSottoCaso])\***

- **IDBase:** codice che, combinato con la *Tipologia*, identifica il requisito generale;
- **IDSottoCaso:** codice progressivo opzionale che identifica gli eventuali sottocasi del requisito.

Ogni requisito riporta lo stato di avanzamento *Completo/Incompleto* raggiunto al termine del progetto di stage.

## Requisiti Funzionali

| Codice | Descrizione                                                                                                             | Stato      |
|--------|-------------------------------------------------------------------------------------------------------------------------|------------|
| ROF1   | L'utente può caricare i dati nell'applicativo della Rete mediante l'uso di file                                         | Completato |
| ROF1.1 | L'utente può caricare i dati nell'applicativo della Rete mediante l'uso di file in formato testo (txt)                  | Completato |
| ROF1.2 | L'utente può caricare i dati nell'applicativo della Rete mediante l'uso di file in formato comma-separated values (csv) | Completato |
| ROF2   | L'utente può effettuare la configurazione dell'architettura direttamente dall'applicativo                               | Completato |
| ROF2.1 | L'utente può effettuare la configurazione dell'architettura mediante l'inserimento di nuovi hidden layers               | Completato |
| ROF2.2 | L'utente può effettuare la configurazione dell'architettura mediante la cancellazione di hidden layers già presenti     | Completato |
| ROF2.3 | L'utente può effettuare la configurazione dell'architettura mediante la modifica di hidden layers già presenti          | Completato |
| ROF2.3 | L'utente per effettuare la configurazione dell'architettura deve inserire almeno un hidden layers                       | Completato |
| ROF3   | L'utente può effettuare del vettore previsione su cui effettuare la previsione sui dati addestrati della Rete           | Completato |
| ROF4   | L'utente visualizza il valore delle previsioni ottenute dalla Rete nella textarea dedicata                              | Completato |
| RDF5   | L'utente può salvare i dati addestrati della Rete                                                                       | Completato |
| RDF6   | L'utente può visualizzare i dati addestrati della Rete salvati in formato JSON                                          | Completato |
| RDF7   | L'utente può visualizzare i dati inseriti della Rete e su cui effettuare la previsione                                  | Completato |
| RDF8   | L'utente può effettuare l'inserimento dei parametri di clusterizzazione sui dati addestrati della Rete                  | Completato |

|         |                                                                                                                                                            |            |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| RDF9.1  | L'utente può effettuare l'inserimento del primo parametro di differenziale di accoppiamento per effettuare la clusterizzazione delle domande               | Completato |
| RDF9.2  | L'utente può effettuare l'inserimento del secondo parametro di differenziale di accoppiamento per effettuare la clusterizzazione delle domande             | Completato |
| RDF9.3  | L'utente può effettuare l'inserimento del secondo parametro di differenziale di accoppiamento per effettuare la clusterizzazione delle domande             | Completato |
| RDF10   | L'utente visualizza mediante tecnologia canvas i risultati di previsione                                                                                   | Completato |
| RDF10.1 | L'utente visualizza mediante tecnologia canvas il dettaglio dei risultati di previsione                                                                    | Completato |
| RDF11   | L'Utente può scegliere il numero di domande, tra quelle proposte, su cui effettuare l'analisi della Rete                                                   | Completato |
| RDF11.1 | L'Utente può scegliere come numero di domande su cui effettuare l'analisi della Rete la dimensione 89                                                      | Completato |
| RDF11.2 | L'Utente può scegliere come numero di domande su cui effettuare l'analisi della Rete la dimensione 120                                                     | Completato |
| RFF12   | L'utente visualizza il valore delle previsioni espresse in termini RGB colours ottenute dalla Rete nella textarea dedicata                                 | Completato |
| RFF13   | L'utente visualizza la clusterizzazione delle domande ottenuta dai parametri di differenziale di accoppiamento ottenute dalla Rete nella textarea dedicata | Completato |
| RFF14   | L'utente visualizza i dati inseriti della Rete all'interno nella textarea dedicata dell'applicativo                                                        | Completato |
| RFF15   | L'utente può visualizzare i dati inseriti nella Rete su una pagina secondaria reindirizzata dall'applicativo                                               | Completato |
| RFF16   | L'utente può effettuare la cancellazione dello storico delle operazioni eseguite                                                                           | Completato |
| RFF17   | L'utente può effettuare la cancellazione dello storico delle operazioni eseguite                                                                           | Completato |

---

Tabella 1: Tabella dei Requisiti Funzionali

### Requisiti di Vincolo

| Codice | Descrizione                                                                                                                            | Stato      |
|--------|----------------------------------------------------------------------------------------------------------------------------------------|------------|
| ROV1   | Deve essere possibile utilizzare l'applicativo almeno in uno dei seguenti browser: Chrome, Firefox, Internet Explorer e Microsoft Edge | Completato |
| RDV2   | La Rete Neurale deve essere sviluppata usando ConvNet.JS                                                                               | Completato |
| RDV3   | il linguaggio consigliato per lo sviluppo è javascript e HTML                                                                          | Completato |
| RDV4   | L'interfaccia grafica deve impedire l'immissione di dati dai box di input che possano corrompere l'integrità della Rete                | Completato |
| RFV5   | L'interfaccia grafica deve essere il più possibile user friendly per l'utente                                                          | Completato |

Tabella 2: Tabella dei Requisiti di Vincolo

### Requisiti di Qualità

| Codice | Descrizione                                                                    | Stato      |
|--------|--------------------------------------------------------------------------------|------------|
| ROQ1   | Redigere la documentazione inerente all'Analisi dei Requisiti per Rete Neurale | Completato |

Tabella 3: Tabella dei Requisiti di Qualità

### Resoconto dei Requisiti

| Tipo              | Obbligatorio | Facoltativo | Desiderabile | Totale    |
|-------------------|--------------|-------------|--------------|-----------|
| <b>Funzionale</b> | 10           | 12          | 6            | <b>28</b> |
| <b>di Vincolo</b> | 1            | 3           | 1            | <b>5</b>  |
| <b>di Qualità</b> | 1            | 0           | 0            | <b>1</b>  |
| <b>Totale</b>     | <b>12</b>    | <b>16</b>   | <b>7</b>     | <b>35</b> |

Tabella 4: Resoconto dei Requisiti

### 2.3.3 Interfaccia utente Rete neurale - Definitiva

#### Questions test - Prevision Neural Net

##### CONFIGURAZIONE RETE NEURALE

```
layer_defs = []
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth: 6});
layer_defs.push({type:'fc', num_neurons:2, activation:'tanh'});
layer_defs.push({type:'fc', num_neurons:2, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:6});

net = new convnetjs.Net();
net.makeLayers(layer_defs);

trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01, momentum:0.1, batch_size:10, l2_decay:0.001});
```

##### OUTPUT DELLA RETE

[create JSON](#) [clean](#)

Richiesta di inserimento:  
 - Parametri di differenziale tra le domande (valore valido  $\geq 0$ ). Utili per creare dei cluster parametrizzati in base al codice del colore  
 - Dati del vettore previsione  
 Successivamente premere il pulsante di previsione

Richiesta di previsione inviata alla rete ...  
 Il vettore [0,0,0,0,0,0] ha previsione calcolata di  
 [0.5062846448317758, 0.3321313045918288, -0.32621589743208046, 0.49428328908691455, 0.3386207175509344, -0.2744086771129914]

Analisi dei dati  
 domanda 1: 125-255-125 - domanda 2: 178-255-170 - domanda 3: 255-171-171 - domanda 4: 128-255-128 - domanda 5: 168-255-168 - domanda 6: 255-185-185  
 Parametri di differenziale dei gruppi, su cui si vuole valutare matematicamente il comportamento della Rete: 0 0 0  
 Relazione fra le domande con un parametri differenziazione configurati  
 cluster 1: 1  
 cluster 2: 2  
 cluster 3: 3  
 cluster 4: 4  
 cluster 5: 5  
 cluster 6: 6



Figura 19: Interfaccia finale utente della Rete neurale di prova - Visione box inserimento architettura e visualizzazione storico.

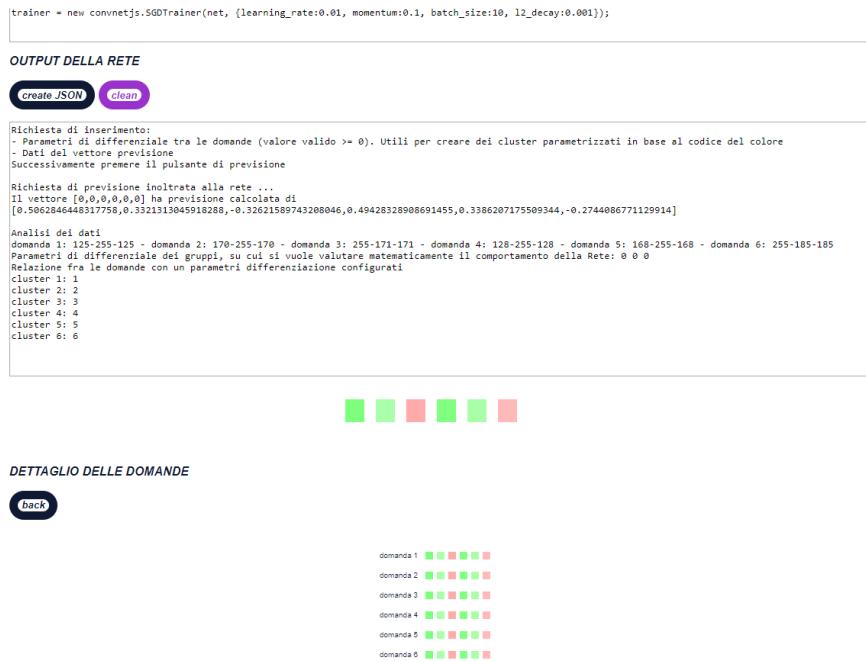


Figura 20: Interfaccia finale utente della Rete neurale di prova - Visione canvas generale e di dettaglio.

**OUTPUT DELLA RETE**

[create JSON](#) [clean](#)

```
Parametri di differenziale dei gruppi, su cui si vuole valutare matematicamente il comportamento della Rete: 8 8 8
Relazione fra le domande con un parametri differenziazione configurati
cluster 1: 1, 4
cluster 2: 2, 5
cluster 3: 3
cluster 4: 6

Richiesta di previsione inoltrata alla rete ...
Il vettore [0,0,0,0,0] ha previsione calcolata di
[0.5062846449317758,0.3321313045918288,-0.32621589743208046,0.49428328908691455,0.3386207175509344,-0.2744086771129914]

Analisi dei dati
domanda 1: 125-255-125 - domanda 2: 170-255-170 - domanda 3: 255-171-171 - domanda 4: 128-255-128 - domanda 5: 168-255-168 - domanda 6: 255-185-185
Parametri di differenziale dei gruppi, su cui si vuole valutare matematicamente il comportamento della Rete: 8 16 16
Relazione fra le domande con un parametri differenziazione configurati
cluster 1: 1, 4
cluster 2: 2, 5
cluster 3: 3, 6
```

[go to detail](#)

**INSERIMENTO DATI DI PREVISIONE**

Differenziale di accoppiamento:

Vettore di previsione

| Domanda 1    | Domanda 2    | Domanda 3    |
|--------------|--------------|--------------|
| -1 ⊕ 0 ⊕ 1 ⊕ | -1 ⊕ 0 ⊕ 1 ⊕ | -1 ⊕ 0 ⊕ 1 ⊕ |

| Domanda 4    | Domanda 5    | Domanda 6    |
|--------------|--------------|--------------|
| -1 ⊕ 0 ⊕ 1 ⊕ | -1 ⊕ 0 ⊕ 1 ⊕ | -1 ⊕ 0 ⊕ 1 ⊕ |

[prevision trainer](#)

Figura 21: Interfaccia finale utente della Rete neurale di prova - Visione inserimento parametri di previsione.

## Questions test - Prevision Neural Net

Applicativo sviluppato per predire le risposte dei candidati durante un colloquio al fine di assunzione

Numero di domande nel database

89 ▾ Confirm

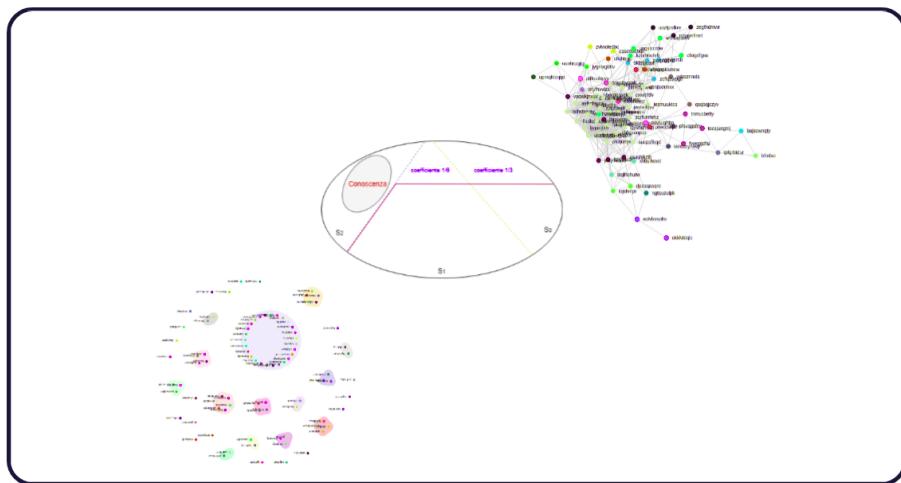


Figura 22: Interfaccia finale utente della Rete neurale del database - Visuale di accesso alla rete, con scelta del numero di domande da analizzare.

### Questions test - Prevision Neural Net

#### CONFIGURAZIONE RETE NEURALE

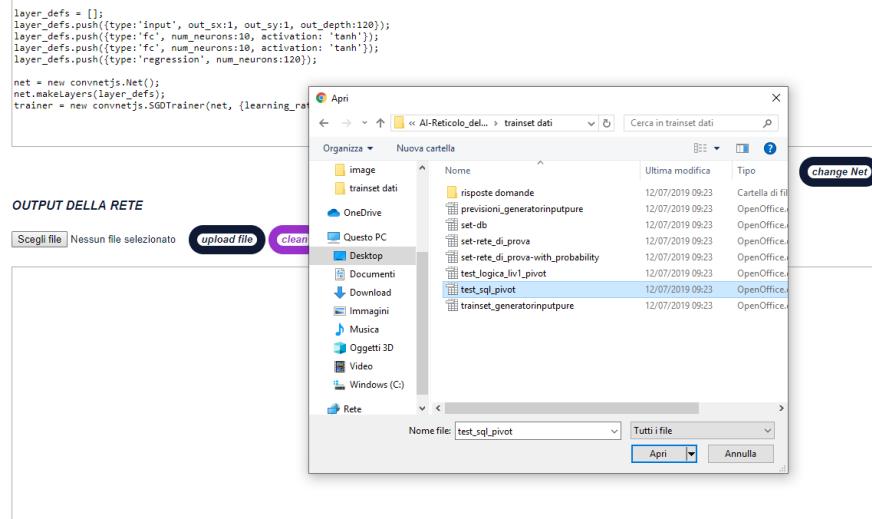


Figura 23: Interfaccia finale utente della Rete neurale del database - Visione caricamento file dati.



Figura 24: Interfaccia finale utente della Rete neurale del database - Visione box con storico delle operazioni e canvas generale.

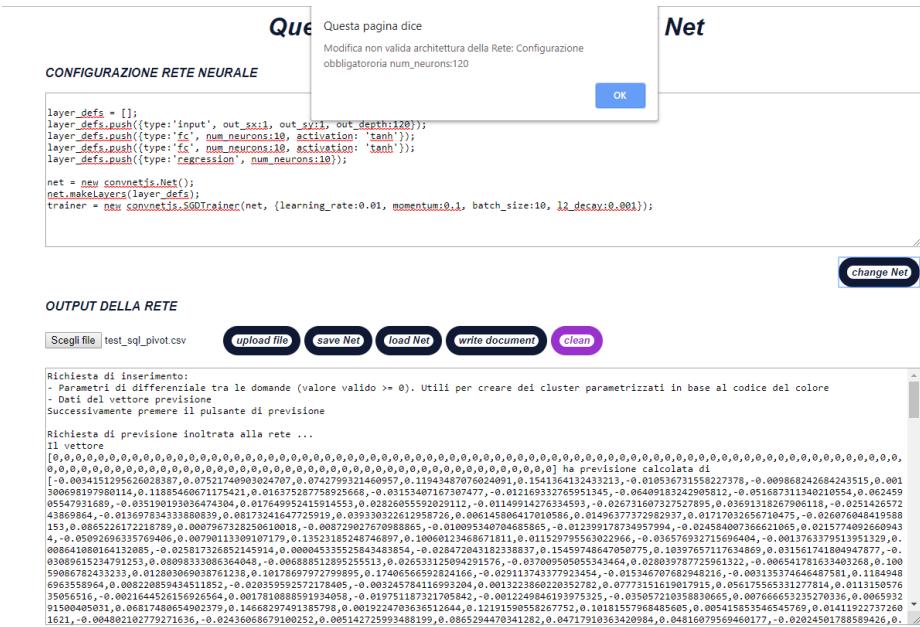


Figura 25: Interfaccia finale utente della Rete neurale del database - Visione errore rilevato durante la configurazione dell'architettura.

## 3 Principal Component Analysis

La **Principal Component Analysis** (acronimo PCA) è una tecnica impiegata nell'ambito della statistica multivariata<sup>3</sup> usata per semplificare i dati d'origine.

Lo scopo che tale tecnica persegue è lo studio della relazione esistenti tra i campioni d'interesse, con la riduzione di un numero più o meno elevato di variabili. La riduzione dimensionale avviene tramite una trasformazione lineare delle variabili coinvolte, con lo scopo di effettuare la proiezione di quelle originarie in un nuovo sistema cartesiano, in cui tutte le variabili vengono ordinate in maniera decrescente per ordine di varianza. Successivamente la variabile con maggiore varianza viene proiettata sul primo asse, la seconda sul secondo e via via sempre così per tutte le variabili sotto esame. La PCA è particolarmente utile quando la dimensionalità dello spazio delle misure è elevata (molte colonne); tuttavia i campioni si vengono a trovare in uno spazio di dimensioni significativamente ridotte.

Indispensabile per la buona riuscita della tecnica è, come già esposto sopra, la ricerca del numero di componenti principali significative, ovvero tutte le variabili coinvolte a meno di quelle legate al rumore. Il "rumore" è sempre concentrato nelle ultime variabili, non includerle nell'analisi dei dati porta a dati più puliti, con un rapporto segnale/rumore più alto.

Il calcolo di quali sono le componenti più significative si ottiene con la varianza. La prima componente principale spiega la massima percentuale di variabilità presente nei dati rappresentabili in una singola dimensione, ciò significa la direzione lungo cui si registra la massima dispersione dei dati (quanto il valore medio si discosta). La varianza porta il vantaggio di essere indipendente dal sistema di riferimento, una rotazione degli assi mantiene immutata la varianza totale all'interno del sistema.

### 3.1 Metodologia applicata

L'analisi e l'implementazione della tecnica di **Principal Component Analysis** l'ho svolta nel periodo 17 - 21 luglio con la seguente metodologia:

- Studio e configurazione di R e R-Studio;
- Studio del significato della tecnica e sua correlazione con le Rete neurali;
- Studio della metodologia di implementazione;
- Scelta del package più adeguato per svolgere l'applicazione della PCA sul set di dati;

---

<sup>3</sup>parte della statistica in cui l'oggetto dell'analisi è almeno composta da due elementi

- Implementazione del metodo della PCA e dei metodi di supporto ad analisi dei dati;
- Analisi dei risultati ottenuti dalla PCA;
- Analisi delle domande presenti nel database aziendale;
- Confronto dei risultati della PCA con quelli ottenuti dalla Rete neurale e con i test di logica presenti nel database.

Ogni implementazione ed osservazione l'ho svolta prima nei dati di prova e solo successivamente gli ho traslati all'interno delle domande nel database. In questo modo ho sempre avuto piena conoscenza della correttezza e delle possibili correlazioni vigenti nel metodo in uso.

### 3.2 Sviluppo

La tecnica di PCA effettua analisi dei dati. I dati che ho impiegato sono i medesimi che ho utilizzato per effettuare il trainset della Rete neurale. Il formato utilizzato è sempre CSV e i dati parsati dal software R sono strutturati per riga con i risultati di ogni test, e per colonna con il risultato di una specifica domanda k.

1. Caricamento del package **factoextra**: la mia scelta è ricaduta su tale package e non in altri con la medesima funzione, per via della visualizzazione dei dati elegante basata su ggplot2;
2. **Caricamento del file CSV** in memoria per mezzo della trasposizione in data frame;
3. **Standardizzazione** dei dati. Tale compito mi è risultato indispensabile perchè anche se di norma, la standardizzazione, viene usata per evitare situazioni erronee (alcune variabili X presentate con una variabilità molto maggiore rispetto ad altre) e io nei casi in analisi ero in possesso di dati già presentati con la medesima scala; avendo la necessità di individuare gli autovettori e applicare il criterio di Keiser nel set di dati, ho dovuto procedere ugualmente a standardizzare.
4. Calcolo della **PCA** con la seguente formula:

```
prcomp(df_numeric, scale = FALSE)
```

R offre due metodi per calcolare la PCA:

- *prcomp(x, scale = FALSE)*: dove *x* rappresenta una matrice numerica o data frame e *scale* un valore logico che indica se le variabili devono essere ridimensionate/standardizzate;

- *princomp(x, cor = FALSE, scores = TRUE)*: dove *x* rappresenta una matrice numerica o data frame, *cor* valore logico che se a true ridimensiona e centra i dati prima di procedere all'analisi e *scores* valore logico che se a true calcola le coordinate su ciascun componente principale.

Ho deciso di far uso della funzione *prcomp* perchè usa la decomposizione del valore singolare (SGV) che offre una precisione leggermente migliore rispetto all'uso del metodo *princomp*. Il metodo *prcomp* include nei propri elementi di output:

- (a) *sdev*: deviazione standard delle componenti principali;
  - (b) *rotation*: la matrice dei carichi delle variabili, ovvero le colonne degli autovettori;
  - (c) *center*: la media variabile, indica se le variabili devono essere spostate per essere centrate sullo zero;
  - (d) *scale*: deviazione standard delle variabili;
  - (e) *x*: coordinate degli individui sulle componenti principali.
5. Calcolo degli **autovalori della matrice di covarianza**. Mostra la percentuale di varianze di competenza di ciascun componente principale con

```
get_eig(res.pca)
```

6. **Riepilogo** mediante il metodo *summary* dei risultati ottenuti dal calcolo della PCA. Effettua la standardizzazione dei risultati ottenuti e ne calcola gli autovalori di covarianza con individuazione di deviazione standard, proporzione della varianza e proporzione cumulata.

```
summary(res.pca)
```

7. **Calcolo degli autovettori** con

```
loadings(res.pca)
```

8. **Individuazione della matrice correlazione** dei dati con

```
cor(df_numeric)
```

9. Creazione di un data frame contenente per ogni variabile principale quali sono le componenti ordinate in modo decrescente strettamente correlate.

### 3.3 Risultati ottenuti

Le immagini relative ai risultati delle domande di logica sono, per motivi di spazio, parziali.

#### 3.3.1 Calcolo della Principal Component Analysis

```
> prcomp(df_numeric, scale = FALSE)
Standard deviations (1, ..., p=6):
[1] 1.5905164 0.9270882 0.6549647 0.2877259 0.2567623 0.2267723

Rotation (n x k) = (6 x 6):
PC1          PC2          PC3          PC4          PC5          PC6
V1 -0.3136368 -0.06351095 0.6506652 -0.0008078537 0.660023581 0.196472324
V2  0.3888043 0.53581153 0.2479470 0.7072146111 -0.008846399 0.005359089
V3 -0.5186977 0.45305827 -0.1486278 0.0011520653 0.150101946 -0.693591199
V4 -0.3009414 -0.07357741 0.6112341 -0.0012775235 -0.719969817 -0.109796655
V5  0.3876316 0.53818077 0.2452635 -0.7069936980 -0.008478550 0.006089356
V6 -0.4905017 0.45668194 -0.2435273 0.0019475491 -0.152746837 0.684257823
```

Figura 26: Visualizzazione del calcolo della PCA tramite prcomop per il set di prova.

```
> prcomp(df_numeric, scale = FALSE)
Standard deviations (1, ..., p=89):
[1] 0.58043495 0.5588746 0.5497946 0.53126069 0.52059133 0.5177118 0.51146770 0.49349655 0.48594238 0.47508962 0.46691672 0.45874610 0.45991145 0.44254212 0.43589886 0.43233385 0.42689299
[19] 0.41870599 0.41659129 0.41384863 0.40256268 0.39119893 0.38976118 0.38548521 0.38101770 0.37484799 0.37051118 0.36871104 0.36502116 0.35881986 0.351352328 0.348913224 0.34691783
[37] 0.34271023 0.33901596 0.32552453 0.33754741 0.32925297 0.32469961 0.32275238 0.31788589 0.31716223 0.31403459 0.30962209 0.30523077 0.30243228 0.29816979 0.29591333 0.29215156 0.28983447 0.28421446
[55] 0.28177829 0.28074860 0.27884498 0.27585038 0.27424743 0.27139654 0.27019256 0.26679640 0.26420824 0.25674716 0.25253431 0.24954843 0.24591523 0.24355894
[73] 0.24187129 0.23789184 0.23498756 0.23280141 0.22959880 0.22384654 0.21341210 0.19833321 0.18912521 0.17449777 0.15236693 0.14016744 0.13282731 0.10140188 0.08865511 0.07203589 0.02534543

Rotation (n x k) = (89 x 89):
PC1          PC2          PC3          PC4          PC5          PC6          PC7          PC8          PC9          PC10         PC11         PC12         PC13         PC14
V1  2.636387e-03 0.0053601056 -4.587466e-03 0.0047285977 -6.267813e-04 0.0012179541 -0.000678271 0.001559397 -1.333849e-03 0.000106171 -0.0037520794 0.0133571069 -0.012466511 0.01669655
V2  1.186062e-01 -0.0983273120 -2.069575e-01 -0.159589648 1.666802e-01 0.5078924212 -0.4156370156 -0.0055475540 2.878968e-01 0.2664054993 -0.214239846 -0.1154210418 -0.004759859
V3  2.418305e-03 -0.0075282317 2.123714e-02 -0.0138787491 3.664982e-03 0.0173104077 0.0139035261 -0.0076250979 1.464138e-02 -0.026843815 0.015897238 0.012222675 0.0332085902 -0.007904186
V4  -6.730258e-03 0.0157451591 -2.178888e-02 -0.0117814649 4.127785e-02 0.0022788841 -0.000299239 -0.034803671 7.185564e-02 0.0206808584 -0.0812277662 -0.0255778432 -0.0022726268 -0.000337898
V5  0.397820e-02 0.0133866499 1.167013e-02 -0.0933261778 2.604523e-05 0.0141769888 -0.0067276127 -0.0041570833 -2.553817e-02 0.0384179571 0.0119898680 -0.0186997858 0.0477993376 0.022818524
V6  -3.397820e-02 0.0133866499 1.167013e-02 -0.0933261778 2.604523e-05 0.0141769888 -0.0067276127 -0.0041570833 -2.553817e-02 0.0384179571 0.0119898680 -0.0186997858 0.0477993376 0.022818524
V7  1.15955e-02 0.0070792227 -0.653489e-02 0.0332109315 4.940894e-02 -0.0281294496 -0.0265328818 -0.0067276127 -0.0035010869 -0.053437983 -0.053437983 -0.053437983 -0.053437983 -0.053437983
V8  6.752564e-02 0.1271216308 -0.687290e-02 0.2670813592 2.783984e-02 -0.0864297064 0.0020403117 0.0012793572 -0.578918e-02 -0.0252448861 0.034861717 -0.0236246302 -0.0258588808 -0.170862744
V9  4.916774e-02 0.0313626989 -0.185083e-02 0.1604337886 1.244642e-01 0.0159574415 0.0160172949 -0.0249788166 -7.438952e-02 0.1016247255 -0.0382499359 0.1526616739 0.044115389
V10 1.070129e-02 0.0154179359 -0.185083e-02 0.1604337886 1.244642e-01 0.0159574415 0.0160172949 -0.0249788166 -7.438952e-02 0.1016247255 -0.0382499359 0.1526616739 0.044115389
V11  0.576331e-02 0.0376436565 -7.111694e-02 0.1541793591 -0.0215191e-02 0.0593646457 0.0593646457 -0.163752238 -0.255992e-01 0.1334747690 0.1520353897 -0.0648031778 0.0998464942 -0.12623030
V12  1.707129e-02 0.161914972 -0.161914972 -0.161914972 0.0471039798 0.0078876045 -0.0847098767 -1.602491e-01 0.0471039798 0.1299999588 0.070102762 -0.070102762 -0.070102762 -0.070102762
V13  2.441796e-04 0.0088153860 -1.048816e-02 -0.0395492127 -4.480856e-02 0.0159669461 0.0252357671 0.0631559225 8.3533141e-03 0.01615540106 0.0464469945 0.0494653204 0.0498919256 0.049566387
```

Figura 27: Visualizzazione del calcolo della PCA tramite prcomop per il set delle domande nel database.

**Osservazioni** Osservando esclusivamente la deviazione standard ottenuta dal set di prova, appare come i PC da prendere in considerazione per l'analisi sono i primi tre. Inoltre nella rotazione appare chiaro che le valutazioni ottenute da V1 sono in relazione con V4, V3 con V6 e V2 con V5. Tuttavia queste sono solo mere osservazioni senza ancora alcuna prova matematica, completa a supporto.

Per quanto concerne il set delle domande nel database è molto difficile fare qualunque tipo di assunzione sulla natura dei dati causa la loro numerosità.

### 3.3.2 Calcolo degli autovettori ed individuazione di Summary

```
> summary(res.pca)
Importance of components:
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6
Standard deviation     1.9297771 1.0581456 0.9099156 0.38288957 0.34464930 0.244856110
Proportion of Variance 0.6209837 0.1867054 0.1380601 0.02444629 0.01980709 0.009997418
Cumulative Proportion 0.6209837 0.8076891 0.9457492 0.97019549 0.99000258 1.000000000
> # fornisce gli autovalori della matrice di covarianza/varianza delle dimensioni principali
> get_eig(res.pca)
eigenvalue variance.percent cumulative.variance.percent
Dim.1 3.72403947   62.0983737          62.09837
Dim.2 1.11967213   18.6705374          80.76891
Dim.3 0.82794633 13.8060084          94.57492
Dim.4 0.14660442  2.4446293          97.01955
Dim.5 0.11878314   1.9807094          99.00026
Dim.6 0.05995451   0.9997418          100.00000
```

Figura 28: Visualizzazione del calcolo del metodo summary ed individuazione degli autovalori per il set di prova.

```
> summary(res.pca)
Importance of components:
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6
Standard deviation     1.43136144 1.30933969 1.27234891 1.25540770 1.24609579 1.22206570
   1.21141257 1.20907948 1.17990883 1.17845300 1.16204099 1.15753299
   1.15663032 1.14998663
Proportion of Variance 0.02103668 0.01972807 0.01820813 0.01772265 0.01746071 0.01698565
   0.01654052 0.01634701 0.01534701 0.01506496 0.01504464 0.01484788
   0.01504464 0.01484788
Cumulative Proportion 0.02393868 0.04231675 0.06695249 0.07824758 0.09578028 0.11269334
   0.12948710 0.14617321 0.16261151 0.17861659 0.19424270 0.22499327
   0.2406328 0.25510814 0.26995692
   0.26995692
   Comp.7    Comp.8    Comp.9    Comp.10   Comp.11   Comp.12   Comp.13   Comp.14   Comp.15   Comp.16
   0.09218745 0.08747001 0.08377745 0.08027500 0.07677400 0.07327300 0.07077300 0.06827300 0.06627300 0.06427300
   0.06427300
   Comp.17   Comp.18   Comp.19   Comp.20   Comp.21   Comp.22   Comp.23   Comp.24   Comp.25   Comp.26
   0.06177805 0.06142451 0.06117382 0.06113382 0.06110382 0.06075370 0.06045370 0.06015370 0.05985370 0.05955370
   0.05955370
   Comp.27   Comp.28   Comp.29   Comp.30   Comp.31   Comp.32
   0.05925370 0.05895370 0.05865370 0.05835370 0.05805370 0.05775370
   0.05775370
   Comp.33   Comp.34   Comp.35   Comp.36   Comp.37   Comp.38   Comp.39   Comp.40   Comp.41   Comp.42
   0.05745370 0.05715370 0.05685370 0.05655370 0.05625370 0.05595370 0.05565370 0.05535370 0.05505370 0.05475370
   0.05475370
   Comp.43   Comp.44   Comp.45   Comp.46   Comp.47   Comp.48
   0.05445370 0.05415370 0.05385370 0.05355370 0.05325370 0.05295370 0.05265370 0.05235370 0.05205370 0.05175370
   0.05175370
   Comp.49   Comp.50   Comp.51   Comp.52   Comp.53   Comp.54   Comp.55   Comp.56   Comp.57   Comp.58
   0.05145370 0.05115370 0.05085370 0.05055370 0.05025370 0.05095370 0.05065370 0.05035370 0.05005370 0.04975370
   0.04975370
   Comp.59   Comp.60   Comp.61   Comp.62   Comp.63
   0.04945370 0.04915370 0.04885370 0.04855370 0.04825370
   0.04825370
   Comp.64   Comp.65   Comp.66   Comp.67   Comp.68   Comp.69   Comp.70   Comp.71   Comp.72   Comp.73
   0.04795370 0.04765370 0.04735370 0.04705370 0.04675370 0.04645370 0.04615370 0.04585370 0.04555370 0.04525370
   0.04525370
   Comp.74   Comp.75   Comp.76   Comp.77   Comp.78
   0.04495370 0.04465370 0.04435370 0.04405370 0.04375370
   0.04375370
   Comp.79   Comp.80   Comp.81   Comp.82   Comp.83   Comp.84   Comp.85   Comp.86   Comp.87   Comp.88
   0.04345370 0.04315370 0.04285370 0.04255370 0.04225370 0.04195370 0.04165370 0.04135370 0.04105370 0.04075370
   0.04075370
   Comp.89
   Standard deviation 0.79836916 0.796146684 0.782386719 0.772551946 0.759737073 0.755173964 0.745553834 0.728101546 0.717568688 0.699402168
   0.667162768
   Proportion of Variance 0.007167148 0.007127429 0.006683383 0.006711419 0.006495611 0.00641863 0.006217062 0.005963326 0.005798608 0.005402106 0.002428461
   0.002377065
   Cumulative Proportion 0.04057395 0.047786081 0.054584963 0.061295482 0.067786093 0.074190956 0.080441018 0.086441018 0.092377344 0.097569319 1.000000000
```

Figura 29: Visualizzazione del calcolo del metodo summary per il set delle domande nel database.

```

> get_eig(res.pca)
      eigenvalue variance.percent cumulative.variance.percent
Dim.1   2.0487956    2.3038680          2.303868
Dim.2   1.7143702    1.9278071          4.231675
Dim.3   1.6192262    1.8208178          6.052493
Dim.4   1.5760487    1.7722647          7.824758
Dim.5   1.5527547    1.7460707          9.570828
Dim.6   1.5104555    1.6985053         11.269334
Dim.7   1.4934446    1.6793765         12.948710
Dim.8   1.4838713    1.6686114         14.617321
Dim.9   1.4618708    1.6438718         16.261193
Dim.10  1.4227291    1.5998571         17.861050
Dim.11  1.3902728    1.5633600         19.424410
Dim.12  1.3699621    1.5405206         20.964931
Dim.13  1.3647870    1.5347012         22.499632
Dim.14  1.3398826    1.5066963         24.006328
Dim.15  1.3379168    1.5044857         25.510814
Dim.16  1.3204001    1.4847882         26.995602
Dim.17  1.2959135    1.4572531         28.452855
Dim.18  1.2743942    1.4330546         29.885910
Dim.19  1.2640485    1.4214209         31.307331
Dim.20  1.2428757    1.3976122         32.704943
Dim.21  1.2240074    1.3763948         34.081338
Dim.22  1.2233920    1.3757027         35.457041
Dim.23  1.2060378    1.3561879         36.813229
Dim.24  1.1911601    1.3394580         38.152687
Dim.25  1.1747105    1.3209604         39.473647
Dim.26  1.1645347    1.3095178         40.783165
Dim.27  1.1600618    1.3044880         42.087653
Dim.28  1.1465801    1.2893278         43.376981
Dim.29  1.1296923    1.2703376         44.647318
Dim.30  1.1166696    1.2556935         45.903012
Dim.31  1.1093668    1.2474815         47.150493
Dim.32  1.1050109    1.2425833         48.393076
Dim.33  1.0885526    1.2240760         49.617152

```

Figura 30: Individuazione degli autovalori per il set delle domande nel database.

**Osservazioni** È essenziale per individuare il numero di componenti (PC) necessarie per effettuare un’analisi corretta, basarsi sul calcolo della *variance.percent* o *Proportion of Variance*. Lo scopo è individuare le componenti principali che catturano la maggior parte di variabilità dei dati.

Nel caso del set di prova sono sufficienti le variabili PC1, PC2 e PC3 che catturano già il 93% della variabilità.

Quanto appena descritto si può riscontrare anche graficamente, come presentato di seguito.

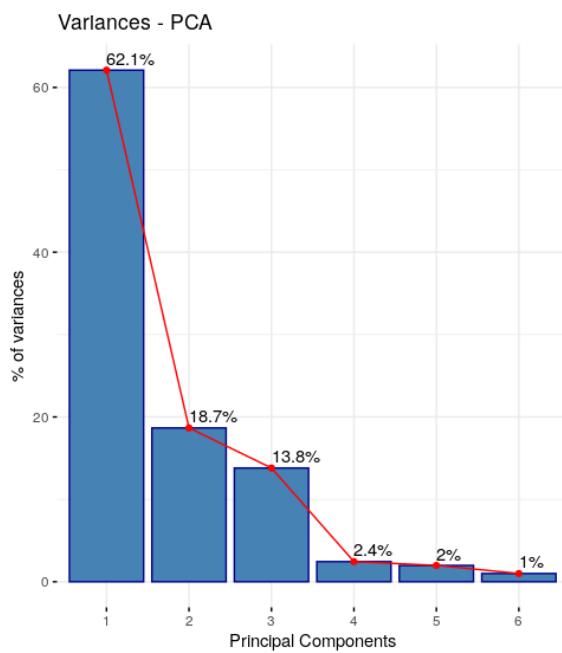


Figura 31: Plot della rappresentazione grafica della varianza sui PC del set di prova.

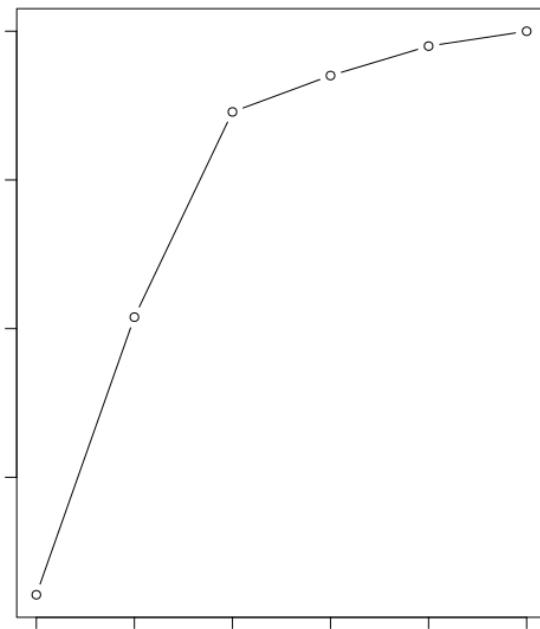


Figura 32: Plot della rappresentazione grafica della varianza sui PC del set di prova.

Le ultime PC4, PC5 e PC6 hanno una variabilità molto bassa, trascurabile.

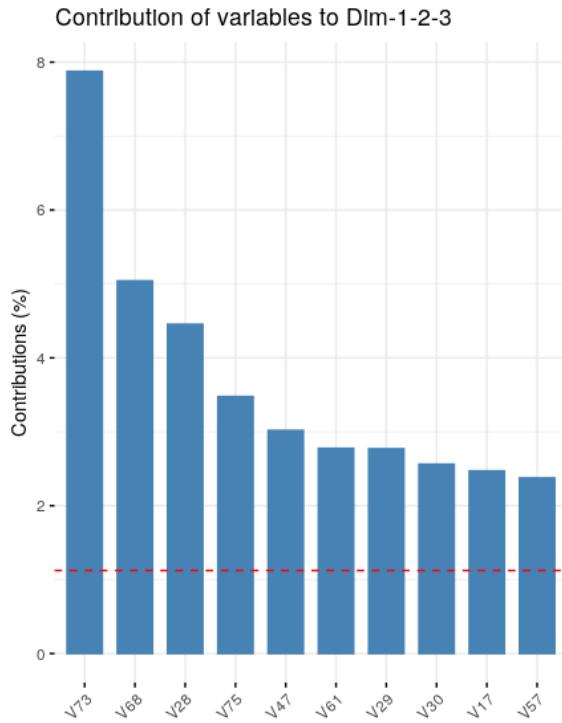


Figura 33: Rappresentazione grafica di come la varianza si distribuisce sulle PC individuate dal modello sul set di prova.

Tuttavia per quanto riguarda le domande nel database ogni conclusione "a occhio" risulta impossibile da effettuare sempre a causa della numerosità dei dati di set. L'utilizzo di un analisi dei risultati per mezzo di plot è l'unica via percorribile.

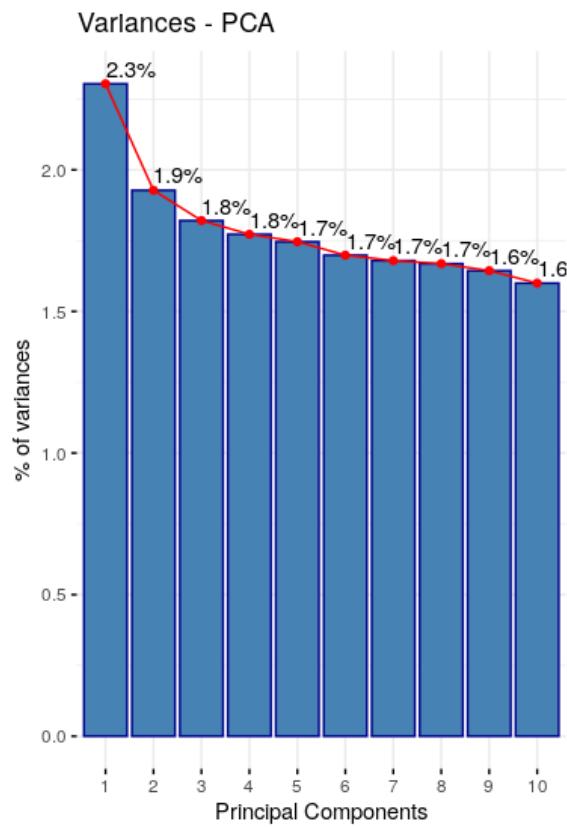


Figura 34: Plot della rappresentazione grafica della varianza dei primi dieci PC del set delle domande nel database.

Il plot in figura 34 mostra le prime dieci componenti; tutte si presentano con una varianza molto bassa. A causa di ciò, per poter affermare quante PC sono indispensabili per una valutazione oggettiva dei dati, è indispensabile avere una visione totalitaria di tutte variabili coinvolte nel modello.

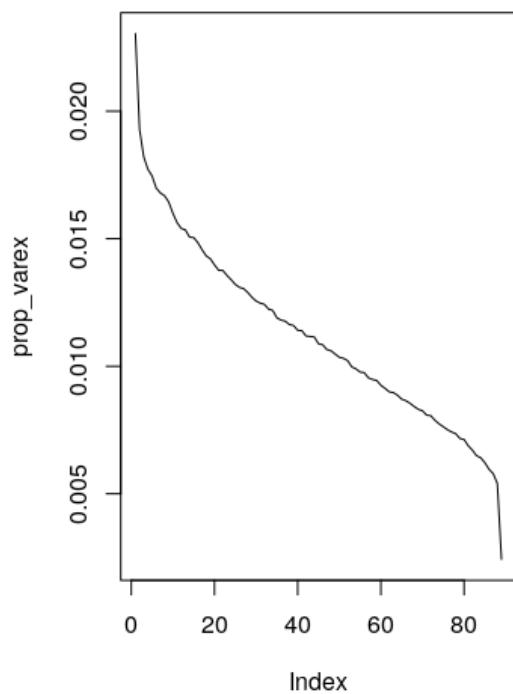


Figura 35: Plot della rappresentazione grafica della varianza di tutti i PC del set delle domande nel database.

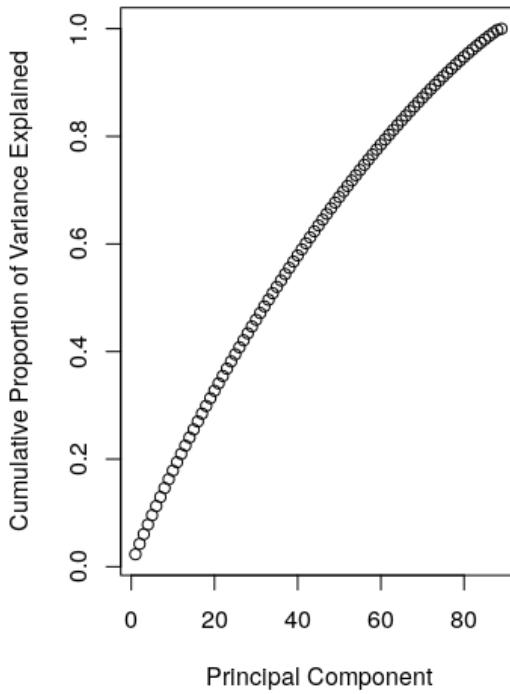


Figura 36: Plot della rappresentazione grafica della varianza di tutti i PC del set delle domande nel database.

Non risulta sufficiente l'andamento dei plot per riuscire a definire il numero adeguato di componenti principali da utilizzare. In questi casi, è buona norma, fare riferimento a tre criteri:

- *Quota della varianza totale*: si deve considerare un numero di PC tale che tenga conto di una percentuale sufficientemente elevata di varianza totale, proporzionale al numero di variabili originarie (ovvero più è alto il numero di componenti del modello e più è accettata una percentuale minore di varianza spiegata).
- *Screen-graph*: fa uso dei plot degli autovalori in funzione al numero di PC. Gli autovalori sono decrescenti, per cui il grafico ha una buona possibilità, di assumere la forma di una spezzata con pendenza negativa.
- *Eigenvalue one o Regola di Kaiser*: afferma di considerare tutte ed esclusivamente le PC con autovalore maggiore di 1.

Per soddisfare il *primo criterio* è sufficiente fare riferimento a *variance.percent* in *get\_eigen* o *Proportion of Variance* in *summary*. Da questa asserzione ne consegue un quesito: quale è il numero di varianza accettabile avendo un numero di variabili molto elevato?. Per poter rispondere risulta indispensabile procedere al soddisfacimento del *terzo criterio*. Gli autovalori di tutte le componenti coinvolte si possono vedere nella *Standard deviation* risultante dalla *summary*. Le PC del modello, con autovalore superiore a 1, sono le PC contenute nell'intervallo 1-41. Il *secondo criterio*, invece, in questo caso specifico non è in grado di dirmi molto; la diminuzione degli autovalori è graduale, senza salti evidenti.

In conclusione, ho considerato come una percentuale di copertura adeguata quella fornita dalle 41 prime componenti principali (PC).

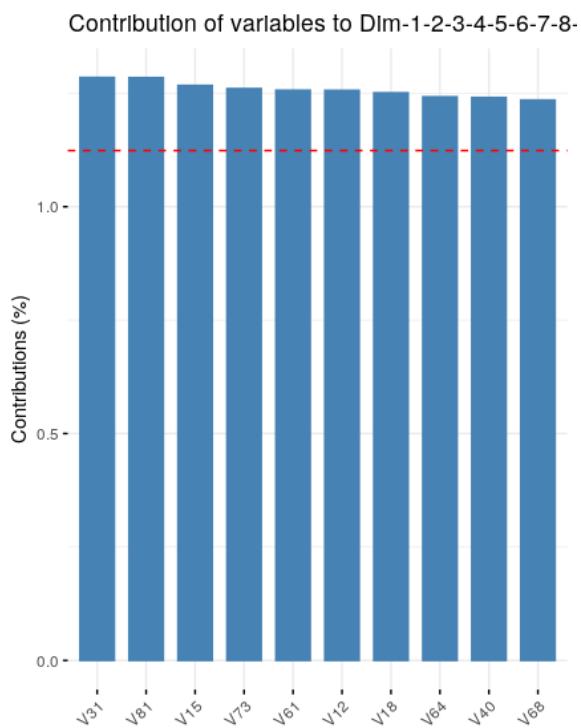


Figura 37: Rappresentazione grafica distribuisce della varianza, sulle PC individuate dal modello, sul set delle domande nel database.

### 3.3.3 Calcolo degli autovettori

```
> loadings(res.pca)

Loadings:
  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
V1  0.420   0.569  0.702
V2 -0.394  0.515  0.282   -0.707
V3  0.418  0.477 -0.278   -0.721
V4  0.421   0.565 -0.709
V5 -0.393  0.518  0.279   0.707
V6  0.402  0.489 -0.350   0.689

  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
SS loadings    1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

Figura 38: Autovettori del set di test.

**Osservazioni** L’analisi dei loadings sulle componenti principali permette di determinare il contributo delle variabili originarie al modello PC.

Per quanto riguarda l’analisi dei dati di prova, la prima variabile ha valori 0.420, 0.569 e 0.702, rispettivamente nelle componenti 1, 3 e 4, legandosi ai risultati della variabile 4 (che riempie le medesime componenti oscillando di poco nei valori presentati). Il medesimo match coinvolge le variabili 2 con 5 e 3 con 6.

Tali assunzioni sono molto più complesse da effettuare per i dati delle domande di logica; per agevolare la situazione ha calcolato la matrice correlazione.

### 3.3.4 Calcolo della matrice di correlazione

```
> cor.pca
      V1        V2        V3        V4        V5        V6
V1  1.000000 -0.4725415  0.5354458  0.8534648 -0.4724833  0.4749618
V2 -0.4725415  1.0000000 -0.4025775 -0.4839185  0.8811501 -0.3892455
V3  0.5354458 -0.4025775  1.0000000  0.5237420 -0.3996431  0.9371832
V4  0.8534648 -0.4839185  0.5237420  1.0000000 -0.4838589  0.4736120
V5 -0.4724833  0.8811501 -0.3996431 -0.4838589  1.0000000 -0.3862259
V6  0.4749618 -0.3892455  0.9371832  0.4736120 -0.3862259  1.0000000
```

Figura 39: Autovettori del set di test.

Figura 40: Autovettori del set delle domande nel database.

I CSV generati di correlazione sono i seguenti:

|   | A                                                                                                                                             |
|---|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | V1" "V2" "V3" "V4" "V5" "V6<br>"1" 1 4 3 6 5 2<br>"2" 2 5 6 3 1 4<br>"3" 3 6 1 4 5 2<br>"4" 4 1 3 6 5 2<br>"5" 5 2 6 3 1 4<br>"6" 6 3 1 4 5 2 |
| 2 |                                                                                                                                               |

Figura 41: CSV generato a partire dalla matrice correlazione del set della rete di prova.

domande di logica.

I file CSV già generati con i seguenti passaggi:

- Creando la matrice correlazione, che mostra il grado di correlazione di ogni componente con ogni variabile del modello;
  - Creando un data frame, che mostra per ogni variabile, quale è la componente a cui si collega in ordine decrescente (da quella che si correla di più a quella che si correla di meno).
  - Salvando il data frame, del punto precedente, in un file .csv;

Da tale distribuzione dei dati, per il set dei dati di prova emerge quello che mi aspettavo. Per esempio la variabile 1 si correla con grado massimo con se stessa, successivamente con la componente 4 (che è la sua domanda sorella) e successivamente con le domande 3 e 6 (i suoi genitori) e alla fine con le domande 2 e 5 (nella realtà queste ultime due non hanno alcuna correlazione con la domanda 1). Tale ragionamento vale per tutte le altre 5 componenti, i cui risultati rimangono coerenti con le aspettative e con ciò che viene dichiarato nel Grafo della Conoscenza, rappresentato nella figura 6 nella sezione § 2.1.

Pensavo di poter effettuare il medesimo ragionamento anche per l'analisi dei dati delle domande di logica, tuttavia questo non è stato possibile. Osservando il file csv generato ho riscontrato i seguenti fenomeni:

- In primo luogo, se una variabile nella matrice correlazione è dichiarata in correlazione con una componente, questa poi non risulta in correlazione con la variabile;
- Analizzando il testo delle domande, ho riscontrato come il modello prodotto crea relazioni strette fra domande appartenenti trivialmente a categorie diverse, (ad esempio domande che trattano relazioni con altre di serie numeriche) e meno con domande che parlano dello stesso tema;
- Dal plot delle variabili viene mostrato come le stesse indicate come correlate risultano, invece, sparse impedendo la formazione dei cluster. Su un numero n grande di variabili (come le 89 domande di logica) tale fenomeno impedisce a chi interpreta il modello, di individuare una correlazione tra il grafico delle variabili e la matrice di correlazione.

A questo punto ho iniziato a farmi delle domande. Escludendo un mio errore di codifica (dopo aver opportunamente effettuato un accurato controllo del codice da me prodotto) e trovandomi di fronte ad una situazione dove i risultati inerenti ai dati di prova soddisfano appieno le attese; ho iniziato a pensare che centrasse la possibilità che ha un candidato di indovinare correttamente una risposta ad una domanda.

Ho rifatto, perciò, tutta l'analisi su un nuovo modello, adoperando i dati di prova generati da n input sottoposti alla probabilità di indovinare. Come viene illustrato dai plots nella sezione seguente, i risultati generati con i dati di prova puro vengono totalmente falsati quando si tiene conto di tale eventualità, rispetto alle aspettative imposte dal Grafo della Conoscenza.

### 3.4 Conclusione dell'analisi

In conclusione, un test con domande a risposta multipla concede ad un candidato l'elevata possibilità di indovinare le domande che non sa; e il modello creato dalla Principal Component Analysis con i dati delle risposte alle domande di logica, ne fornisce la prova.

Se la possibilità di indovinare fosse bassa i risultati del modello in esame ne verrebbero appena "sporcati", e invece invalidano completamente la possibilità di avere un risultato attendibile e coerente con la realtà.

Le figure seguenti mostrano proprio tale fenomeno: nel caso dell'indovinato le domande con una maggiore correlazione non creano alcun cluster.

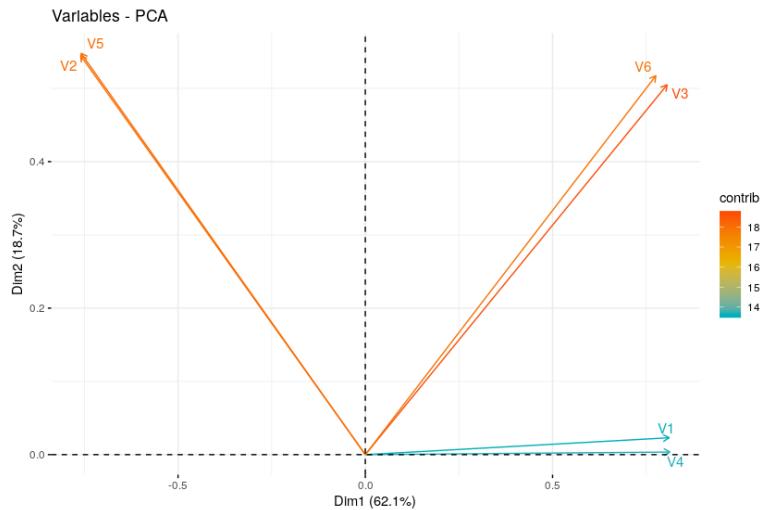


Figura 43: Rappresentazione per mezzo di plot di come le variabili si presentano nelle due componenti principali con il calcolo della PCA - utilizzo del set di prova puro sul grafo della conoscenza.

```
> correlazione
 [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]    1    4    3    6    5    2
 [2,]    2    5    6    3    1    4
 [3,]    3    6    1    4    5    2
 [4,]    4    1    3    6    5    2
 [5,]    5    2    6    3    1    4
 [6,]    6    3    1    4    5    2
```

Figura 44: Le domande in correlazione con la variabili principali in ordine decrescente sui dati di prova.

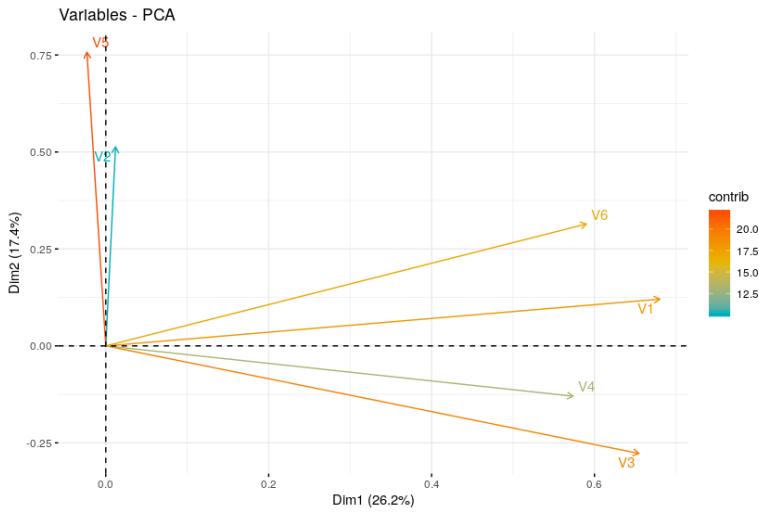


Figura 45: Rappresentazione per mezzo di plot di come le variabili si presentano nelle due componenti principali con il calcolo della PCA - utilizzo del set di prova spurio con la probabilità di indovinare.

```
> correlazione
[,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    6    3    4    5    2
[2,]    2    6    5    4    1    3
[3,]    3    1    4    6    2    5
[4,]    4    3    6    1    2    5
[5,]    5    6    2    1    4    3
[6,]    6    1    4    3    5    2
```

Figura 46: Le domande in correlazione con le variabili principali in ordine decrescente sui dati di prova soggetti alla probabilità di indovinare.

L'analisi del modello ha evidenziato come il Reticolo della Conoscenza costruito dai dati delle domande nel database, una volta generato, risentirà inevitabilmente del fattore di rischio "domande indovinate". Mi aspetto, per cui, che in rapporto al contenuto delle stesse, gli argomenti e la difficoltà delle domande appartenenti al medesimo cluster possano non essere coincidenti.

Per completezza di seguito è presentato il plot generato dal calcolo della PCA, applicato alle variabili dei dati di logica.

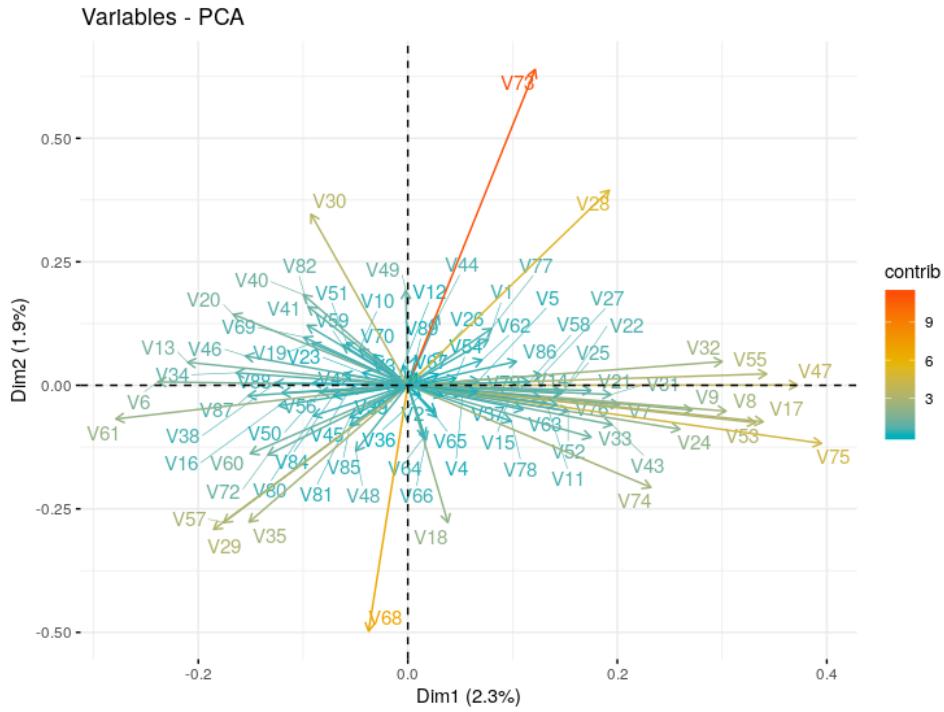


Figura 47: Le domande in correlazione con le variabili principali in ordine decrescente sui dati di prova soggetti alla probabilità di indovinare.

L'unica possibilità per continuare la costruzione del Reticolo è rivolgere nuovamente l'attenzione alla Rete neurale, studiandone i cluster possibili per mezzo delle previsioni assunte da ogni variabile, sulla base della previsione standard<sup>4</sup>.

---

<sup>4</sup>Vettore previsione tutto a zero

## 4 Costruzione del Reticolo della Conoscenza

Dal 26/07 al 28/07 ho iniziato la costruzione del Reticolo della Conoscenza. Per realizzarlo ho utilizzato un applicativo sviluppato durante un precedente stage dell'azienda, che fa largo uso della libreria d3.js. Svolge elaborazione di dati, presentandoli in *Cluster Based* o *Force Based*, a discrezione delle esigenze dell'utente.

L'attività di creazione, documentazione e test del Reticolo è durata fino al termine dello stage.

### 4.1 Descrizione del sistema

L'applicazione accetta in input file di estensione CSV. Ogni colonna di quest'ultimo viene interpretata dal sistema come un parametro da elaborare; per questo la prima riga del file deve essere o preceduta da una dichiarazione di variabili, tante quante sono le colonne da parametrizzare, oppure è la prima riga di dati che viene interpretata come una dichiarazione e conseguentemente non rappresentata all'interno del Reticolo.

Nel sistema possono venire settati i seguenti aspetti:

- La *tipologia* di Reticolo:
  - Cluster Based: raggruppa un insieme di oggetti in modo tale che tutti gli elementi contenuti nel medesimo cluster sono più simili l'uno all'altro rispetto a quelli contenuti in altri gruppi;
  - Force Based: in base alla forza di ogni nodo, viene rappresentata come un'unica regione compatta le istanze appartenenti alla medesima classe, in cui vengono visivamente identificati i percorsi di differenziazione. Nel layout le celle differenzianti sono poste in prossimità della classe più fortemente correlata.
- *Normalizzazione* dei dati in input:
  - No: non viene applicata alcuna tecnica di normalizzazione dei dati;
  - MinMax: i dati vengono ridimensionati su un intervallo specifico ( $\min$ ,  $\max$ ), tuttavia tale tecnica non è in grado di gestire i valori anomali;
  - Gaussian: o normale in cui i dati vengono normalizzati in una curva, in cui i valori della stessa grandezza sono soggetti ad approssimazione;
  - Interquartile: si occupa di standardizzare i dati in modo da quantificare l'estensione del 50% della distribuzione del carattere, che si trovano attorno alla mediana;

- Tipologia di *distanza* applicabili ai punti:
  - Euclidea: tiene conto della distanza tra i punti;
  - Camberra: tiene conto della distanza tra le coppie in uno spazio vettoriale;
  - Pearson: distanza di correlazione che misura il grado di correlazione tra due punti. Valuta la covarianza tra due variabili in rapporto al prodotto della deviazione standard. Non è vantaggiosa su dati semplici.
- *Metodo* di associazione dei punti:
  - Single: ”vicino al prossimo”, la distanza fra i gruppi è posta al pari della più piccola delle distanze calcolabili a due a due tra tutti gli elementi del gruppo. Accentua tutte le somiglianze tra i gruppi a discapito della loro differenziazione netta.
  - Complete: ”vicino più lontano”, viene considerata la maggiore tra le distanze calcolate a due a due tra gli elementi di due gruppi. Privilegia la differenziazione tra i gruppi a discapito dell’omogeneità degli elementi in essi contenuti. In questo caso i punti vengono rappresentati come meno compatti e diluiti.
  - Average: viene considerata come distanza fra due gruppi la media fra tutte le distanze calcolate a due a due tra gli elementi dei due gruppi. I risultati ottenibili sono i più attendibili (essendo basato sulla media delle distanze), i gruppi risultano più omogenei e differenziati tra di loro.

Un’ulteriore funzionalità, dell’applicativo, permette all’utente di decidere se si desidera procedere con una rappresentazione del Reticolo manuale, per mezzo di una threshold, o automatica progressiva.

#### 4.1.1 Configurazione

Analizzando l’applicativo in base al carattere dei miei dati in ingresso e alle aspettative sull’output del modello, ho valutato come configurazione più idonea per la formazione del Reticolo della Conoscenza la seguente:

- *Redistance*: No;
- *Normalize*: No;
- *Distance-Type*: Euclidea;
- *Method*: Single.

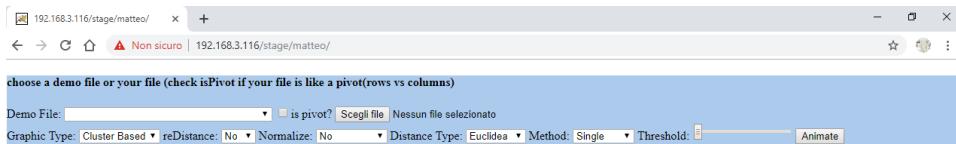


Figura 48: Configurazione usata nel sistema per generare il Reticolo della Conoscenza.

## 4.2 Creazione dei file CSV

Come già accentato all'interno della sezione §4.1 prima di procedere alla creazione del Reticolo, ho dovuto preparare i dati di previsione. Questo è stato reso più agevole grazie alla creazione, da parte mia, di un metodo che ha il compito, una volta messa in funzione la Rete, di calcolare:

1. Le previsioni ottenibili da un vettore di previsione settato a 1 e -1 per ogni singolo elemento;
2. Sui dati del punto (1) un vettore delle differenze dove viene calcolato il delta in rapporto al vettore standard <sup>5</sup>.

Il vettore delle differenze viene stampato su console del browser <sup>6</sup>, in modo che ne basti prelevare il contenuto e inserirlo su un file CSV. Ogni elemento per poter funzionare all'interno dell'applicativo deve essere separato da un ; e ogni riga di previsione deve essere preceduta dal codice della domande in modo da rendere più agevole l'interpretazione del Reticolo. È a discrezione dell'utente l'inserimento di un'ulteriore riga di dichiarazione dei parametri.

## 4.3 Creazione del Reticolo della Conoscenza per sui dati di Prova

### 4.3.1 L'architettura della Rete

Durante la costruzione del Reticolo della Conoscenza è stato importante capire se effettivamente l'architettura scelta in precedenza fosse corretta, e il suo vero significato. La rete sviluppata per il progetto è una Rete neurale statica a multistrato, nella quale sono presenti i seguenti elementi:

- input layer: dove viene dichiarata la dimensione dell'input: lunghezza, larghezza e profondità dei dati da analizzare;

---

<sup>5</sup>Vettore tutto a zero

<sup>6</sup>unica alternativa facendo uso di solo codice javascript

- hidden layer (o layer intermedio): composto da uno o più strati nascosti (o intermedi) costituito da  $m$  neuroni;
  - output layer: costituito da  $p$  neuroni pari al numero di output desiderati.

Una Rete neurale permette di formulare complessi modelli matematici (sistemi di equazioni differenziali ecc.) e di individuarne possibili soluzioni. Le soluzioni si determinano sempre a partire dai dati; tuttavia la scelta del modello<sup>7</sup> è un'azione molto complessa.

L'architettura di Rete ipotizzata per i dati di prova e coerente con quanto affermato nella sezione 2.1.3, è il seguente:

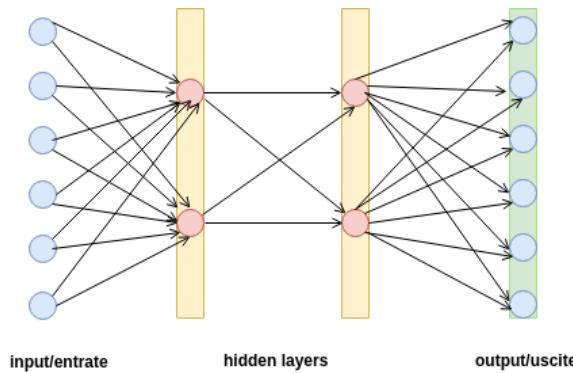


Figura 49: Architettura delle Rete di prova: 6 ingressi in una rete a 3 strati con 2 stati nascosti e 6 uscite.

**4.3.2 Reticolo della Conoscenza generato mediante previsioni sul set di dati che conto esclusivamente della conoscenza del candidato**

Figura 50: file CSV generato per la creazione del Reticolo della Conoscenza sui dati di prova.

Di seguito sono riportate le sequenze di creazione del Reticolo della Conoscenza per i dati di prova che non risentono della probabilità di indovinare di un candidato.

<sup>7</sup>inteso come architettura.



Figura 51: Reticolo della Conoscenza per i dati di prova - Cluster based.

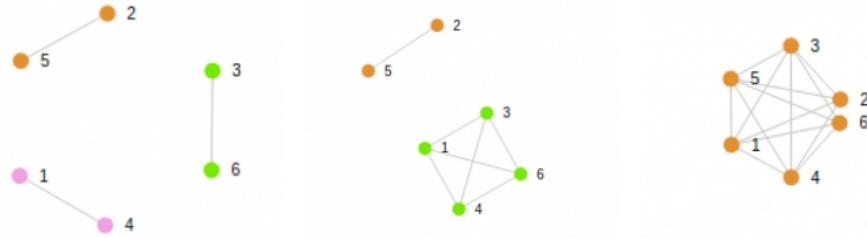


Figura 52: Reticolo della Conoscenza per i dati di prova - Forced based.

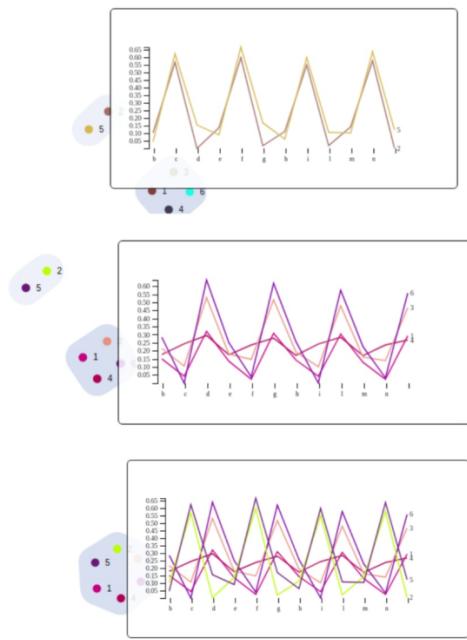


Figura 53: Reticolo della Conoscenza per i dati di prova - Statistiche.

Appare come il Reticolo generato rispetta quanto definito dalla figura 6. Tuttavia effettuando delle ulteriori prove con file dati differenti; ma provenienti dalla medesima Rete neurale, ho riscontrato situazioni contrastanti. L'errore osservato nei casi non corretti riguarda le coppie di domande 1, 4 e 3, 6.

La spiegazione del Reticolo malformato è da ricondurre alla natura stessa dei vettori di input, usati per l'addestramento della Rete di prova. Difatti i valori delle domande correlate sono i medesimi, solo se entrambe le domande sono state poste in fase di test, altrimenti quando la domanda è non posta viene valutata 0. Questo provoca un'oscillazione degli accoppiamenti genitori-figli. Tuttavia è solo uno scostamento causato dall'occorrenza con il quale le domande vengono poste, e di cui l'addestramento della Rete ne risente, come illustrato dalle figure seguenti.

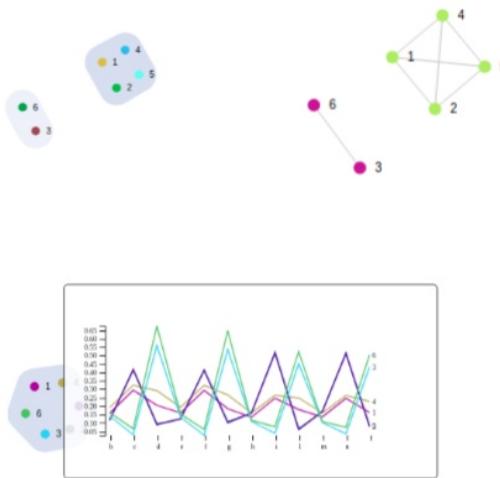


Figura 54: Reticolo della Conoscenza per i dati di prova - Presenza di valori di alterazione.

Le domande 6, 3 sono state poste un numero superiore di volte rispetto alle domande figlie 1, 4, per cui il sistema associa le coppie 1, 4 con le coppie 2, 5 che hanno tra loro una frequenza, dei valori iniziali, più vicina. Inoltre vi è da dire che il metodo Single usato dall'applicativo per associare le domande, nel momento in cui correla tra loro due punti, calcola il prossimo dal complessivo, invece che dal punto con una distanza appena superiore. Per valutare al meglio l'impatto di quanto affermato sopra, in fase di test mi sono preoccupata di analizzare come si comporta la Rete neurale, relativa ai dati di prova, considerando esclusivamente la possibilità che un candidato o risponda correttamente ad una domanda (1) o la sbagli (-1). I risultati di previsione ottenuti sono i seguenti:



Figura 55: Risultati della Rete di prova con un set di dati puro (esclusivamente 1 e -1).

Come si può vedere dalla figura 55 le previsioni ottenute per le coppie di domande (1;4), (3;6) e (2;5) sono identiche, ad eccezioni di alcune variazioni impercettibili da ricondurre ad oscillazioni di Rete; e che non hanno alcun impatto sui risultati da raggiungere. Impostando di ottenere una previsione sul valore 0 e avendo allenato la Rete esclusivamente con valori -1 e 1 mi aspetto che venga effettuata la media, ed è quello che viene fatto effettivamente, per tutte le domande coinvolte. Il fenomeno sorprendente, che mette in luce l'importanza di avere un valore 0 di risposta (non data) nel trainset, è mostrato nelle figure 56 e 57 sotto.

```

layer_defs.push({type:'regression', num_neurons:6});
net = new convnetjs.Net();
net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01, momentum:0.1, batch_size:10, l2_decay:0.001});

```

**OUTPUT DELLA RETE**

[create JSON](#) [clear](#)

Richiesta di inserimento:  
 - Parametri di differenziale tra le domande (valore valido  $\geq 0$ ). Utili per creare dei cluster parametrizzati in base al codice del colore  
 - Dati del vettore previsione  
 Successivamente premere il pulsante di previsione

Richiesta di previsione inoltrata alla rete ...  
 Il vettore  $[-1, 0, 0, 0, 0, 0]$  ha previsione calcolata di  
 $[-0.6797836085310568, 0.4067289774785281, -1.0535753270960428, -0.679783608659422, 0.40672897778163697, -1.0535753269169634]$

Analisi dei dati  
 domanda 1: 255-81 - domanda 2: 151-255-151 - domanda 3: 255-0-0 - domanda 4: 255-81-81 - domanda 5: 151-255-151 - domanda 6: 255-0-0  
 Parametri di differenziale dei gruppi, su cui si vuole valutare matematicamente il comportamento della Rete: 0 0 0  
 Prezzi: fra le domande con un parametri differenziazione configurati  
 cluster 1: 1, 4  
 cluster 2: 2, 5  
 cluster 3: 3, 6

[go to detail](#)

**INSERIMENTO DATI DI PREVISIONE**

Differenziale di accoppiamento:

**Vettore di previsione**

|                           |                           |                           |
|---------------------------|---------------------------|---------------------------|
| Domanda 1<br>-1 Ⓛ 0 Ⓛ 1 Ⓛ | Domanda 2<br>-1 Ⓛ 0 Ⓛ 1 Ⓛ | Domanda 3<br>-1 Ⓛ 0 Ⓛ 1 Ⓛ |
| Domanda 4<br>-1 Ⓛ 0 Ⓛ 1 Ⓛ | Domanda 5<br>-1 Ⓛ 0 Ⓛ 1 Ⓛ | Domanda 6<br>-1 Ⓛ 0 Ⓛ 1 Ⓛ |

[prevision trainer](#)

Figura 56: Risultati della Rete di test con un set di dati puro (esclusivamente 1 e -1), vettore previsione  $[-1, 0, 0, 0, 0, 0]$ .

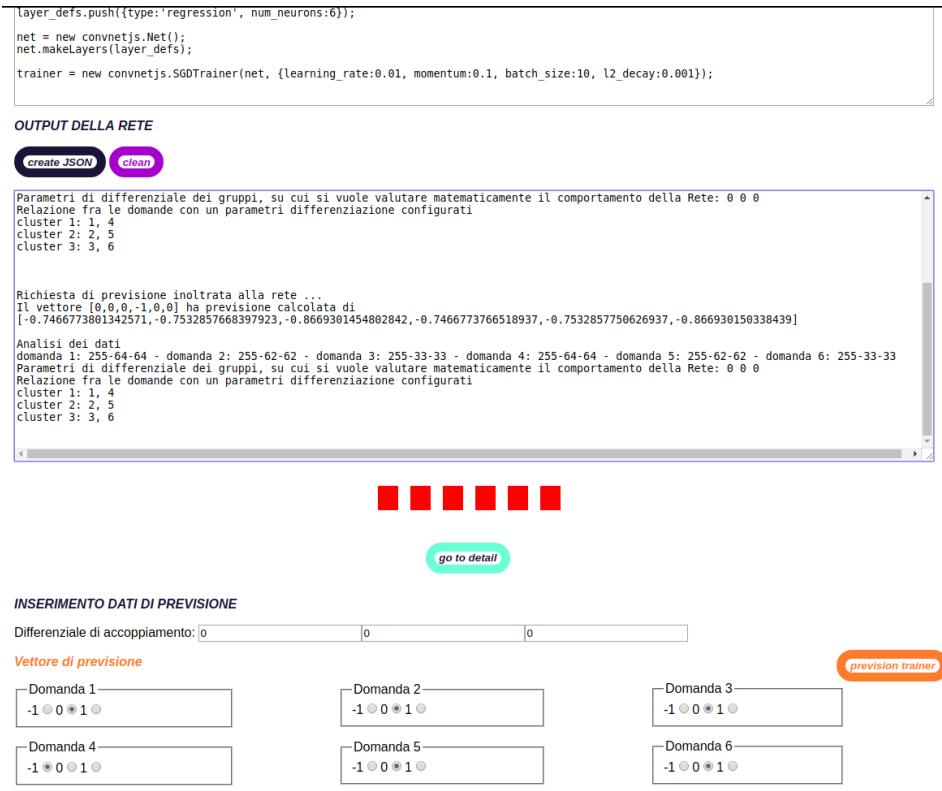


Figura 57: Risultati della Rete di test con un set di dati puro (esclusivamente 1 e -1), vettore previsione  $[0, 0, 0, -1, 0, 0]$ .

Non avere un valore a 0 porta le Rete a valutare come uniche relazioni, strette, esclusivamente quella tra la coppia coinvolta; rendendo meno possibile ogni altra assunzione sulle altre coppie<sup>8</sup>. Ad esempio, le domande 2 e 4 vengono colorate di verdino nella previsione della domanda 1 e di rosso nella domanda 4. Questo ha un impatto negativo sia sulla previsione visiva della Rete, dal quale risulta impossibile la previsione di domande figlie e genitori, che nella costruzione del Reticolo della Conoscenza, in cui non può essere preso come valore di riferimento distanza 0.

La matrice correlazione ottenuta, tuttavia dal modello della PCA, risulta perfettamente allineata con le aspettative delle coppie di domande e delle previsioni, questo è possibile solo perché i dati e le previsioni vengono valutati nella loro generalità e non tra coppia di domande.

<sup>8</sup>La presenza dello 0 mitiga questo aspetto.

### 4.3.3 Reticolo della Conoscenza generato mediante la frequenza calcolata sul set di dati che tiene conto esclusivamente della conoscenza del candidato

Ho ritenuto adeguato studiare il comportamento del set di dati anche in base alla frequenza; in modo da poterne valutare le differenze con la Rete neurale e quali siano, se esistenti, i benefici.

Come primo passo ho provveduto a calcolarmi, per ogni domanda, quale fosse la frequenza delle altre domande, valutate in accordo e in disaccordo con la risposta data. Ho effettuato il procedimento sia nel caso di una domanda con risposta correttamente che non.

```

    * (6: [Array(6), Array(6), Array(6), Array(6), Array(6), Array(6)])  insertData_1a:22
  > 0: (0: ["FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 1: (0: ["FC: 0 1D: 0", "FC: 0 1D: -1", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 2: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 3: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 1"])
  > 4: (0: ["FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 5: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0"])
  > 6: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0"])

  > (6: [Array(6), Array(6), Array(6), Array(6), Array(6), Array(6)])  insertData_1a:23
  > 0: (0: ["FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 1: (0: ["FC: 0 1D: 0", "FC: 0 1D: -1", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 2: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 3: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 0", "FC: 0 1D: 1", "FC: 0 1D: 1"])
  > 4: (0: ["FC: 0 1D: 0", "FC: 0 1D: 1"])
  > 5: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0"])
  > 6: (0: ["FC: 0 1D: 0", "FC: 0 1D: 0"])

```

Figura 58: Frequenza delle domande dei dati di test.

L’immagine sopra illustra, per ogni domanda valutata a 1 (primo set di dati) e a -1 (secondo set di dati), la frequenza che intercorre in tutte le domande con segno coerente o opposto a quello della domanda in esame. Il set di dati viene generato randomicamente con guida<sup>9</sup>. Analizzando i risultati di frequenza ottenuti, ho riscontrato i seguenti aspetti:

- Considerando i valori per riga:
  - per le domande in cui le risposte si presentano a 1, i valori di frequenza che non rispettano le aspettative sono i seguenti:
    - \* la domanda 3 con una frequenza correlata che coinvolge le coppie (1 FC: 0.98, 3 FC: 1) e (4 FC: 0.95, 6 FC: 0.98);
    - \* la domanda 6 con una frequenza correlata che coinvolge le coppie (1 FC: 0.98, 6 FC: 1) e (4 FC: 0.96, 3 FC: 0.95).
  - per le domande in cui le risposte si presentano a -1 i valori di frequenza che non rispetta le aspettative sono i seguenti:
    - \* la domanda 1 con una frequenza correlata che coinvolge le coppie (1 FC:1, 3 FC:1) e (4 FC:0.55, 6 FC:0.80);
    - \* la domanda 4 con una frequenza correlata che coinvolge le coppie (4 FC:1, 3 FC:1) e (1 FC:0.54, 6 FC:0.80);
- Considerando i valori per colonna:
  - la domanda 1 si correla con la domanda 4;

<sup>9</sup>con impiego del Grafo della Conoscenza (mostrato in figura 6)

- la domanda 3 si correla con la domanda 6,
- la domanda 2 con la domanda 5.

Tali considerazioni non sono esclusive per il singolo set di dati; ma rimangono costanti per ogni set di dati generato (anche se di natura parzialmente randomica). Ho riscontrato per ciascuna variabile al massimo un oscillazione che si attesta mai superiore a 0.2.

I valori di frequenza possono concorrere alla formazione del Reticolo della Conoscenza. Per renderlo possibile ho provveduto a creare un file csv, che contenesse per ognuna delle domande, tutte le domande coinvolte valutate con frequenze positive e negative correlate e opposte. La tecnica permette la generazione, per ogni domanda, di un vettore contenente un quantità di dati 4 volte la dimensione dell'input/output della Rete. Il divario permette, positivamente, di poter costruire un Reticolo ove i dati per ogni nodo hanno una stabilità concreta.

Tuttavia il Reticolo generato riscontra delle peculiarità:

- la frequenza non effettua alcuna riduzione dimensionale provocando un utilizzo di spazio pari al numero di domande/dati da analizzare. Ciò comporta che per grandi moli di dati lo spazio necessario cresce in maniera elevata rispetto al numero di elementi di input;
- la frequenza funziona male sui valori anomali.

Il divario tra i valori per riga e colonna sono legittimati dalla natura del set di valori di input. In esso sono contenuti valori 0, e non esclusivamente -1 e 1, causa diretta dell'oscillazione delle correlazioni strette esistenti tra le domande accoppiate <sup>10</sup>.

Il Reticolo generato rispetta fedelmente le aspettative, definite dal Grafo della Conoscenza in figura 6. Tuttavia questo non va valutato come un risultato positivo, in quanto indicatore che ai valori 0, che dovrebbero alterare, anche se marginalmente il risultato finale viene data la medesima importanza non tenendo conto del numero di occorrenze (dimostrazione di come la frequenza gestisce male i valori anomali - vedi esempio sotto). Il beneficio dell'uso della Rete sta proprio in quest'ultimo fattore, con l'utilizzo dell'apprendimento vengono rilevate situazioni che la sola frequenza non è in grado di pesare.

---

Esempio che mette in rapporto la frequenza con le previsioni di una Rete neurale

---

<sup>10</sup>per i dati di prova le domande accoppiate sono (1;4), (3;6) e (2;5)

**Analisi della risposta positiva (FC):**

- (1)  $1/1 = 1$ , 1 risposta positiva tra tutte le domande ( $=1$ ), che sono accoppiate con la domanda  $i$ , quando viene risposta correttamente
- (2)  $8/8 = 1$ , 8 risposte positive tra tutte le domande ( $=8$ ), che sono accoppiate con la domanda  $i + 1$ , quando viene risposta correttamente

- Per la frequenza:  $n$  e  $n + 1$  hanno il medesimo valore di FC
  - Per una Rete neurale: un evento che si verifica 1 volta e un evento che si verifica 8 volte hanno un peso differente, questo perché una rete effettua previsioni non su dati statistici, ma per mezzo di train.
- 

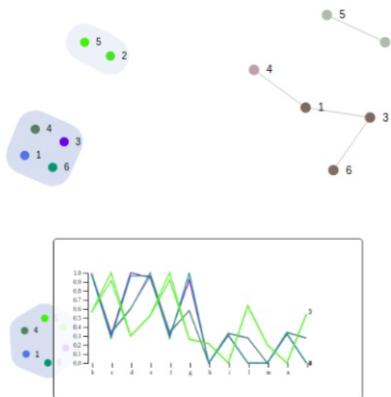


Figura 59: Reticolo della Conoscenza per i dati di prova - Frequenza.

#### 4.3.4 Osservazioni sul Reticolo generato dalle previsioni della rete

In conclusione il Reticolo della Conoscenza, costruito per mezzo delle previsioni della Rete neurale, anche se non sempre perfettamente coerente con le aspettative; nella maggioranza dei casi ricalca fedelmente quanto evidenziato nella figura 6 nella sezione § 2.1.

Quando il Reticolo non è soggetto a deviazioni, gli accoppiamenti tra i punti, rispettano quanto dichiarato dalla matrice correlazione ottenuta dal modello generato dall'uso della PCA, e visibile dall'immagine 41. È obbligo precisare, tuttavia che i risultati della PCA non devono essere presi come assioma assoluto. Le argomentazioni sono le seguenti:

- La Rete neurale, può cogliere oscillazioni che il modello matematico non è in grado, e questo va ad invalidare i risultati ottenuti non solo dalla matrice correlazione, ma anche dell'individuazione dei punti sulle prime due componenti;
- La PCA, come si è visto accadere con la frequenza, non tiene conto del numero di volte cui si verifica un evento, attribuendone il medesimo peso. Invece la Rete, neurale mediante apprendimento, riesce a valutare al meglio, tali fenomeni, avvicinandosi il più possibile alla realtà.
- Lo scopo principale della PCA è quello di spostare gli assi di rappresentazione degli eventi; in modo che vi sia una maggiore facilità di comprensione dei dati, come si ha con la standardizzazione dei valori.

#### **4.3.5 Reticolo della Conoscenza generato con set di dati che tiene conto della possibilità che un candidato ha di indovinare la risposta**

Sia che si tratti di una valutazione per frequenza, che di previsioni ottenute dalla Rete neurale, il Reticolo tende ad ogni nuovo set di dati, ad assumere una diversa conformazione, che ne rende difficile trarne delle conclusioni. In linea generale, si può affermare che:

- Le domande figlie e le domande genitori vengono sempre raggruppate in un singolo insieme, prima di venire fuse in un unico cluster, assieme alle domande con cui non esiste correlazione;
- Il Reticolo della Conoscenza generato con le previsioni della Rete, rispetta meglio le oscillazioni introdotte dalla possibilità di indovinare. Difatti, come vale per il valore 0, la frequenza si dimostra poco capace di cogliere le discrepanze dal modello, non riuscendo a dare il giusto peso ai valori anomali.

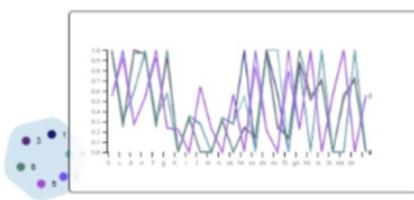
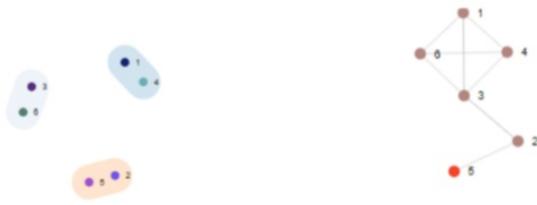


Figura 60: Reticolo della Conoscenza per i dati di prova con indovinato - Frequenza.

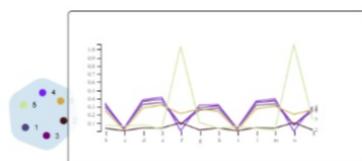


Figura 61: Reticolo della Conoscenza per i dati di prova con indovinato - Previsioni.

#### 4.4 Creazione del Reticolo della Conoscenza per le domande di logica nel database

Requisiti da assolvere, per costruire un Reticolo della Conoscenza valido per le domande contenute nel database aziendale:

1. Studio del testo delle domande per individuare quali sono i gruppi di argomenti;

2. Formulazione di una buona architettura, avvalendosi della grandezza del set di dati, del numero di domande e della conoscenza di quanto emerso dal punto (1);
3. Validazione dei risultati ottenuti, confrontando i cluster con gli argomenti identificati al punto (1). Si deve tenere conto che non si potrà mai ottenere un match perfetto, in quanto il set risente della probabilità di indovinare.

#### 4.4.1 Individuazione dei gruppi di argomenti

Le domande di logica si compongono di due macro-argomenti principali:

- Domande sulle Serie numeriche;
- Domande su diagrammi di Eulero-Venn.

Le restanti domande non presentano un argomento specifico, risultando complesse da catalogare. Queste ultime le ho valutate come punti isolati.

Di seguito, in forma tabellare viene riportata la categorizzazione delle domande.

| Codice      | Argomento   |
|-------------|-------------|
| aekshwnzj   | Nessuno     |
| bajsownqty  | Nessuno     |
| bdwjgxtdsm  | Nessuno     |
| bfmxudzq    | Nessuno     |
| bkfstjjeaw  | Nessuno     |
| bojihohutw  | Nessuno     |
| ccsofpywab  | Eulero-Venn |
| cfkjhtcjjp  | Nessuno     |
| cfxigdfgsu  | Eulero-Venn |
| cidpdlimnu  | Nessuno     |
| cxjganqls   | Nessuno     |
| cxwrjrlnfv  | Nessuno     |
| diucioyoxo  | Nessuno     |
| dpzaqasqro  | Nessuno     |
| ectvbesohs  | Nessuno     |
| efbcntiwr   | Nessuno     |
| elbottcdse  | Nessuno     |
| fbekkymdja  | Nessuno     |
| fclxhnnodej | Nessuno     |
| flsdkdhwpz  | Nessuno     |
| fpiajgnnf   | Nessuno     |
| fupxesqveq  | Nessuno     |
| fvwcpjdfai  | Nessuno     |

|             |                 |
|-------------|-----------------|
| gilviisrcb  | Eulero-Venn     |
| gxushjkzth  | Nessuno         |
| hacwlyvymr  | Nessuno         |
| heohdfuqpi  | Nessuno         |
| heusbnnnto  | Nessuno         |
| hxsywnmxdd  | Nessuno         |
| iesmuuktcs  | Nessuno         |
| ifrcopoblv  | Nessuno         |
| iojsfvnpii  | Nessuno         |
| islhstwuay  | Nessuno         |
| jgcxaxgrrw  | Nessuno         |
| jnxymdjdxl  | Nessuno         |
| jomkrqdwwr  | Nessuno         |
| jwvreibjvv  | Nessuno         |
| jygnlagbhv  | Nessuno         |
| kkhcbynnavp | Nessuno         |
| kprsxjkubz  | Serie numeriche |
| kzorhoohrb  | Nessuno         |
| lhuxxlqtyz  | Nessuno         |
| ljypeljfkhh | Eulero-Venn     |
| lqyvhqfmqd  | Nessuno         |
| lrqrblchp   | Nessuno         |
| mfycpmjyzc  | Nessuno         |
| mhaifmlmrt  | Eulero-Venn     |
| mnrnuysydx  | Serie numeriche |
| ngtquzulpk  | Nessuno         |
| nrawevuhkv  | Nessuno         |
| oaszeadbqh  | Nessuno         |
| obwqtgpzek  | Eulero-Venn     |
| okklvldqlc  | Nessuno         |
| osvluiqhbhp | Nessuno         |
| pfctjunpxs  | Serie numeriche |
| pfsqvgsfen  | Serie numeriche |
| poiuytrewq  | Eulero-Venn     |
| ptfkuubppy  | Nessuno         |
| ptwccskjcf  | Serie numeriche |
| pxvpfrjqnb  | Serie numeriche |
| qbnjsxmnxz  | Serie numeriche |
| qfxndoiusk  | Nessuno         |
| qmzlhlizqng | Nessuno         |
| qsqbqjczyv  | Eulero-Venn     |
| ribhqompqp  | Nessuno         |
| sucpsfikqd  | Nessuno         |
| tnmucbefiy  | Nessuno         |

|              |                 |
|--------------|-----------------|
| tocts juegmj | Nessuno         |
| tpfgrbtdur   | Nessuno         |
| ukafbvkjxl   | Nessuno         |
| ungslccrdw   | Serie numeriche |
| uoohruqjig   | Serie numeriche |
| uoztprdkre   | Nessuno         |
| upmqfdzqqd   | Nessuno         |
| voqwkjnxuv   | Eulero-Venn     |
| vtjiowgfp    | Nessuno         |
| wfthuqobvv   | Nessuno         |
| wpfhnhktzw   | Nessuno         |
| wqkhvmdpol   | Serie numeriche |
| xmqopkamew   | Serie numeriche |
| xnyfhvvlxz   | Serie numeriche |
| xvbulkwvol   | Nessuno         |
| yxkoznnxis   | Nessuno         |
| zcflqcwpgh   | Nessuno         |
| zegfnrndwr   | Nessuno         |
| zptjzdwlf    | Nessuno         |
| zqrfurehmz   | Serie numeriche |
| zrehhcvjyl   | Serie numeriche |
| zvlwoledbc   | Eulero-Venn     |

Tabella 5: Gruppi di argomento per le domande di logica

#### 4.4.2 Architettura della Rete testata

Come già visto all'interno di §2, la scelta dell'architettura adatta per una rete che deve trattare una grande mole di dati e di cui non si ha piena conoscenza del risultato finale perché soggetti all'indovinato come visto in §3 rende tale compito molto arduo.

L'unico modo che ho individuato per affrontare il problema, è procedere con l'analisi del set di dati in input aiutandomi con le conoscenze acquisite durante la costruzione del Reticolo sul set di prova.

Ho testato gli effetti sulla Rete e sulla generazione del Reticolo, di una molteplicità di architetture. Le più rilevanti sono le seguenti:

- `layer_defs = [];`
- `layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});`
- `layer_defs.push({type:'fc', num_neurons:12, activation: 'tanh'});`
- `layer_defs.push({type:'fc', num_neurons:12, activation: 'tanh'});`
- `layer_defs.push({type:'regression', num_neurons:89});`

```

net = new convnetjs.Net();
net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```

Per formulare la configurazione sopra espressa, ho preso come riferimento l'architettura della Rete di prova e ne ho modificato il numero di neuroni per layers, sulla base del aumento del numero di domande e sulla diminuzione del numero di entry nel set di dati di logica. Non posso essere attesi risultati eccelsi, in quanto non vengono considerati il numero di cluser possibili, ma è una buona base di partenza per le osservazioni.

- layer\_defs = [];

```

layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons: 4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:10, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

```
- layer\_defs = [];

```

net = new convnetjs.Net();
net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```
- layer\_defs = [];

```

layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons: 2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:10, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

```
- layer\_defs = [];

```

net = new convnetjs.Net();
net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```
- layer\_defs = [];

```

layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:89});
layer_defs.push({type:'fc', num_neurons: 4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 4, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 5, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 5, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:89});

```
- net = new convnetjs.Net();

```

net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

```

### Prima architettura: layers(12, 12)

Il Reticolo della Conoscenza generato è il seguente:

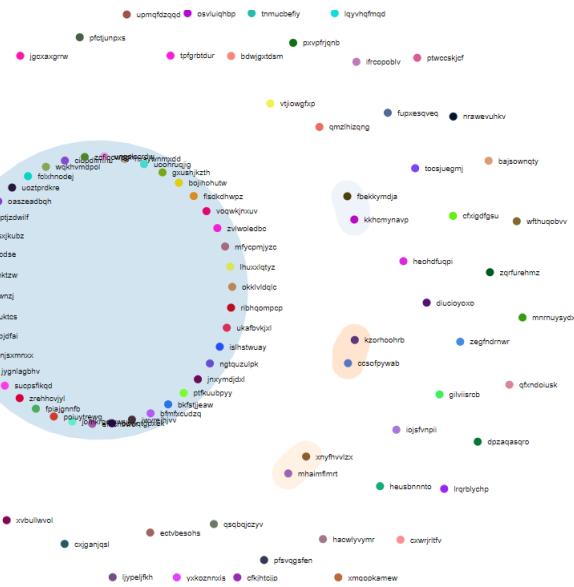


Figura 62: Reticolo della Conoscenza layers(12, 12) - Cluster Based.



Figura 63: Reticolo della Conoscenza layers(12, 12) - Forced Based.

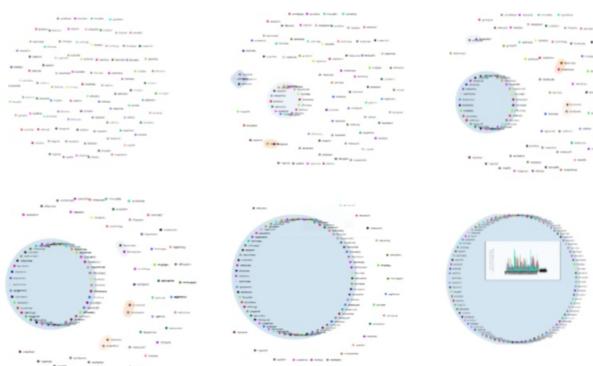


Figura 64: Sequenza di creazione del Reticolo della Conoscenza layers(12, 12) - Cluster Based.

**Osservazioni dei risultati** Prima di tutto si crea un unico cluster, e non due come era atteso, però si vanno ad insinuare un buon numero delle domande sulle Serie numeriche e su Eulero Venn. Tuttavia le domande dello stesso argomento non sono le prime a correlarsi.

| Serie numeriche         | Eulero Venn              |
|-------------------------|--------------------------|
| kprsxjkubz<br>qbhjsxmnx | obwqtgphek<br>poiuytrewq |

|            |            |
|------------|------------|
| ungslccrdw | voqwkjnxuv |
| uoohruqjig | zvlwoledbc |
| wqkhvmdpol |            |
| zrehhcjvl  |            |

Tabella 6: Domande contenute nel primo cluster

Le domande correlate nei cluster (*mhaimflmrt*, *xnyfhvvlxz*), (*ccssofpywab*, *kzorhoohrb*) e (*fbekkymdja*, *kkhcmynavp*), visibili in tabella 62 non presentano alcuna analogia tematica.

Riducendo la correlazione tra i cluster si generano 2 cluster, come visibili dalla tabella 64; ma nuovamente non viene rispettata la differenziazione tra domande Serie numeriche e Eulero Venn.

A seguito di numerosi test che sono succeduti a questa prima configurazione, ho potuto confermare quanto avevo già intuito in precedenza, e descritto in §2.2.2: un numero superiore di 11 neuroni genera una situazione di overfitting delle previsioni della Rete.

### Seconda architettura: layers(4, 4, 10)

Il Reticolo della Conoscenza generato è il seguente:



Figura 65: Reticolo della Conoscenza layers(4, 4, 10) - Cluster Based.

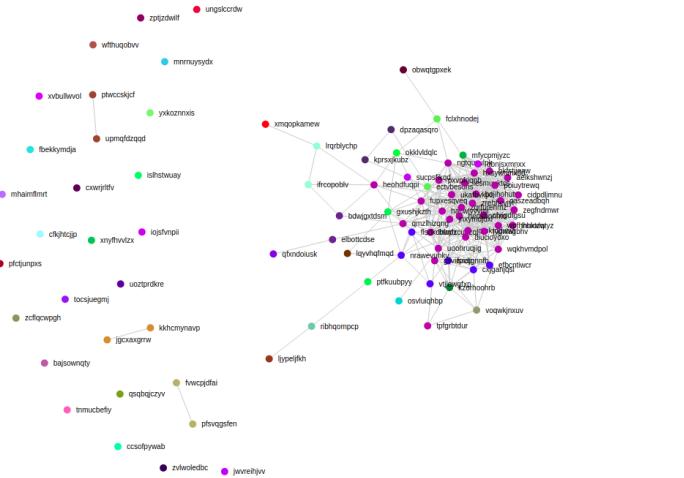


Figura 66: Reticolo della Conoscenza layers(4, 4, 10) - Forced Based.

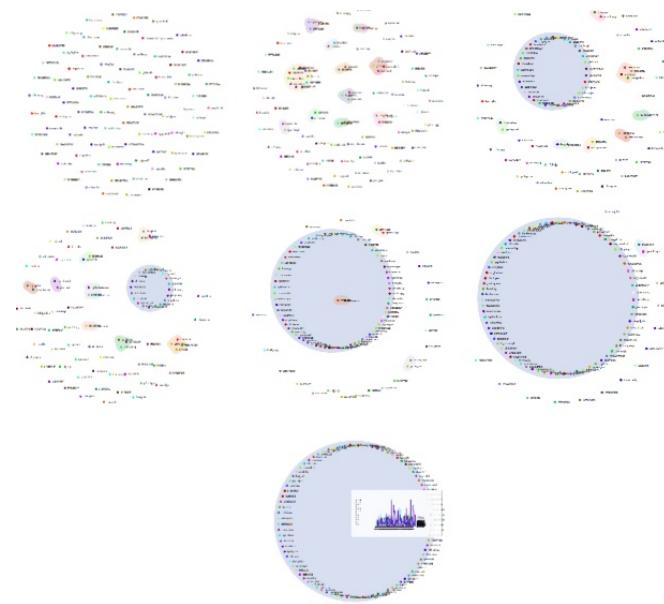


Figura 67: Sequenza di creazione del Reticolo della Conoscenza layers(4, 4, 10) - Cluster Based.

**Osservazioni dei risultati** Questo, ha mio avviso è, uno dei Reticolo più interessanti che ho generato, sulle domande di logica. Esiste un cluster

principale e una decina di gruppi di dimensione via via sempre minore. A mano a mano che viene incrementata la threshold, i gruppi vengono inglobati dal gruppo principale. I punti isolati sono in numero notevole, come si può vedere dalla figura 65 anche se la threeshold di competenza è già quasi a metà del suo percorso.

Questi punti isolati rappresentano, nella maggioranza dei casi, tutte quelle domande che non trovano correlazione né tra le serie numeriche, né tra i diagrammi di Eulero-Venn.

Nel cluster principale vengono posizionati i seguenti punti:

| Serie Numeriche | Eulero-Venn              |
|-----------------|--------------------------|
| zrehhcvjyl      | cfxigdfgsu<br>poiuytrewq |

Tabella 7: Domande contenute nel primo cluster (blu)

Ci sono tuttavia anche altri cluster interessanti:

| Serie Numeriche | Eulero-Venn               |
|-----------------|---------------------------|
| wqkhmdpol       | gilviisrcb<br>voqwkijnxuv |

Tabella 8: Domande contenute nel secondo cluster (rosetta)

| Serie Numeriche                          | Eulero-Venn |
|------------------------------------------|-------------|
| zqrtfurehmz<br>uoohruqjiq<br>zqrtfurehmz |             |

Tabella 9: Domande contenute nel terzo cluster (blu chiaro)

Il primo e il secondo cluster hanno una predominanza delle domande, per quanto concerne le domande catalogate, su Eulero-Venn; e il terzo per le domande sulle Serie numeriche.

I tre gruppi molto probabilmente hanno difficoltà superiori per un candidato e per questo vengono inglobati uno dentro l'altro passo passo. Prima di tutto il primo cluster, influenzato dai dati di probabilità, ingloba il terzo cluster e successivamente il secondo (la domanda *zqrtfurehmz* per un candidato non allenato può risultare più complessa rispetto alla domanda *zrehhcvjyl*).

### Terza architettura: layers(2, 4, 10)

I risultati sono di poco discostanti, rispetto ai risultati ottenuti dalla confi-

gurazione a layers (4, 4, 10). Di seguito riporto esclusivamente la generazione del Reticolo.



Figura 68: Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.



Figura 69: Reticolo della Conoscenza layers(2, 4, 10) - Forced Based.

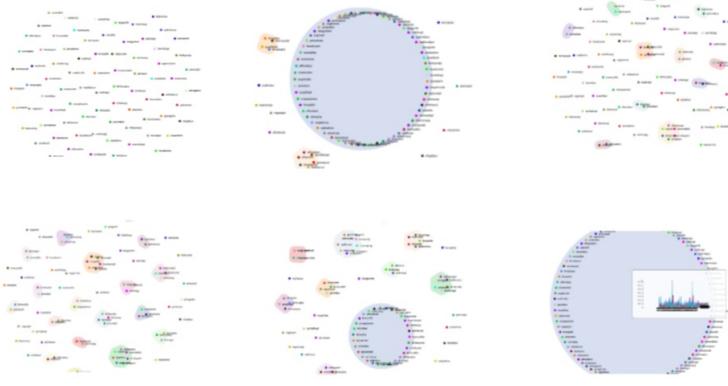


Figura 70: Sequenza di creazione del Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.

#### Quarta architettura: layers(4, 4, 5, 5)

Il Reticolo della Conoscenza generato è il seguente:



Figura 71: Reticolo della Conoscenza layers(4, 4, 5, 5) - Cluster Based.



Figura 72: Reticolo della Conoscenza layers(4, 4, 5, 5) - Forced Based.

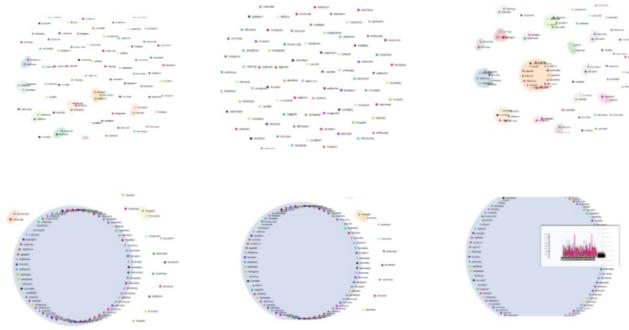


Figura 73: Sequenza di creazione del Reticolo della Conoscenza layers(4, 4, 5, 5) - Cluster Based.

### Osservazioni dei risultati

**Osservazioni dei risultati** La seguente architettura ottiene dei risultati discreti.

Come si può vedere dalle immagini si ha la presenza di più cluster, uno di dimensione maggiore che progressivamente, con l'incremento della threshold ingloba tutti gli altri gruppi. Tuttavia tale configurazione non vede una distinzione netta di argomento per cluster, però al interno le domande relative alle Serie numeriche e ai diagrammi di Eulero-Venn si mostrano strettamente correlate. Di seguito descrivo solo alcuni cluster che ritengo significativi.

| Serie numeriche                        | Eulero Venn              |
|----------------------------------------|--------------------------|
| pfctjunpxs<br>zqrfurehmz<br>wqkhvmdpol | ljypeljfkh<br>poiuytrewq |

Tabella 10: Domande contenute nel primo cluster

Con riferimento alla figura 75 *poiuytrewq* e *ljypeljfhk* derivano dal medesimo cluster, separati dalle domande sulle serie. Come si può vedere dalle immagini in figura 73, alcuni cluster di dimensione ridotta, contengono mini gruppi di domande sulle serie:

- *zrehhcvjyl* e *xnyfhwlzx*; *pxvpfvjqnb* e *qbnjsxnxx*.

Sono presenti anche gruppi isolati che mostrano una relazione strettissima tra domande di Serie numerica e di Eulero-Venn.

- *zcflqcwpgh e gilviisrcb;*
  - *ccsofpywab e xmqopkamew.*

#### 4.4.3 Risultati di frequenza sulle domande di logica

Per maggiore completezza riporto il Reticolo della Conoscenza generato sulle risposte di frequenza, ai test di logica.



Figura 74: Reticolo della Conoscenza layers(4, 4, 10) con Frequenza - Cluster Based.

La situazione non varia molto rispetto a quanto già incontrato per la Rete di prova. La frequenza tende a smorzare le oscillazioni introdotte dalla probabilità e dalla presenza di risposte non date.

Sono individuabili due cluster principali, composti:

| <b>cluster 2 (blu scuro)</b> | <b>cluster 3 (blu chiaro)</b> |
|------------------------------|-------------------------------|
| xmqopkamew                   | gilviisrcb                    |
| zqrfurehmz                   | ccsofpywab                    |
| ptwccskjcf                   | ljypeljfk                     |

Tabella 11: Domande su Serie numeriche e Eulero-Venn contenute nei cluster

Possono apparire come dei buoni risultati; ma siamo ormai consci dell'impossibilità del metodo statistico di effettuare un calcolo pesato delle risposte alle domande. Tale considerazione non fa altro che invalidare l'uso di questo modello. Inoltre, altra nota dolente, le domande individuate sono molto poche, prima di fondersi in un unico cluster. Invece con l'uso dei dati di previsione, è possibile fare assunzioni sulla difficoltà delle domande (grazie alla presenza di un numero molteplice di cluster contenenti domande delle categorie in esame).

#### 4.5 Creazione del Reticolo della Conoscenza per delle domande di SQL nel database

Al termine dello studio e della generazione del Reticolo della Conoscenza per le domande di logica, ho provveduto a generare il Reticolo anche per le domande SQL, per avere una comprensione maggiore di come il Reticolo si comporta con una variazione del trainset.

Le architetture scelte hanno tenuto conto di quanto emerso durante la costruzione dei Reticoli della Conoscenza per le domande di logica.

##### 4.5.1 Individuazione dei gruppi di argomenti

Le domande SQL, sono domande strutturate nel quale risulta complesso individuare quali siano i macro argomenti; tuttavia nei test presi in esame sono presenti, erroneamente, anche delle domande di Informatica Generale. È dunque lecito ipotizzare che le domande di Informatica Generale comprendano dei punti isolati, che saranno gli ultimi a fondersi all'interno del cluster. Di seguito, in forma tabellare viene riportata la categorizzazione delle domande.

| Codice      | Argomento            |
|-------------|----------------------|
| aaxgbiwfen  | SQL                  |
| adiueedxey  | SQL                  |
| afakgklfop  | SQL                  |
| akczmgrdbq  | SQL                  |
| asxurodfne  | SQL                  |
| avagvdjdfu  | SQL                  |
| axrvccwwsp  | SQL                  |
| bbsurwinbt  | SQL                  |
| biweemhojf  | SQL                  |
| blhjkyufeg  | SQL                  |
| bwodxllzmk  | SQL                  |
| bwodxllzmk  | SQL                  |
| csttpqoysn  | SQL                  |
| culepaxyst  | SQL                  |
| cuyddcrymb  | SQL                  |
| cvnhqpgfyp  | SQL                  |
| dagputfvei  | SQL                  |
| dbuqwjsyvj  | SQL                  |
| dkihxiceua  | SQL                  |
| dkzgokqmjq  | SQL                  |
| dlfyuvcbqx  | SQL                  |
| dllpxhzfklt | SQL                  |
| dzxsvydlth  | SQL                  |
| ebrpcrn cwd | SQL                  |
| eftrkdthng  | SQL                  |
| elqzybkpj   | SQL                  |
| enitorndkz  | Informatica Generale |
| enmkdvzmyk  | SQL                  |
| ewplljreqn  | Informatica Generale |
| ffbcwwtnzk  | SQL                  |
| fhqcltuqw   | SQL                  |
| fjmpjvualm  | SQL                  |
| fkdgytnwzf  | SQL                  |
| fumfugvkrl  | Informatica Generale |
| fuzfbwlutx  | SQL                  |
| fyrpicjqan  | SQL                  |
| fzoxjumtyn  | SQL                  |
| gtxwazgyqz  | SQL                  |
| guatssqajf  | SQL                  |
| gyerxpmgqa  | SQL                  |
| hntvpqnrc   | Informatica Generale |
| hyfoktjlpp  | SQL                  |

|            |                      |
|------------|----------------------|
| idfxvfzprt | SQL                  |
| ieimtqtzpk | SQL                  |
| iezewxpthv | SQL                  |
| ifrvxoopny | SQL                  |
| ifvdeyoyre | SQL                  |
| ikyqsdhuit | SQL                  |
| jbpjxhaxyy | SQL                  |
| jedvsptntk | SQL                  |
| jycipstmyp | SQL                  |
| kbnwlxqtqc | SQL                  |
| kexorxwipq | SQL                  |
| kfbpwmrkgz | SQL                  |
| kjecbikedc | SQL                  |
| kqkjrygfsc | SQL                  |
| kzzvvwvnxn | SQL                  |
| lgiqayejyf | SQL                  |
| ljozpygnua | SQL                  |
| lohspalpqn | SQL                  |
| lzpqvyslrv | SQL                  |
| mdwdlluz   | SQL                  |
| mezjrkwdmi | SQL                  |
| mligaxxxyv | SQL                  |
| mljwybuamb | SQL                  |
| naywhaxocj | SQL                  |
| nqrcokzyep | SQL                  |
| oexeosrlxq | Informatica Generale |
| oldnuetsfx | SQL                  |
| ownmhihmlf | SQL                  |
| oxinyiwtdw | SQL                  |
| pdegmbefyz | Informatica Generale |
| pnornaipi  | Informatica Generale |
| pozlktnbwi | Informatica Generale |
| pralrcttdk | Informatica Generale |
| ptfyswhhoi | SQL                  |
| ptvnijvlyr | SQL                  |
| qfpxivqgfs | SQL                  |
| qhpmywunln | SQL                  |
| qhurfkbsbi | SQL                  |
| qyfsksfpqr | SQL                  |
| rcvjbkxijb | Informatica Generale |
| rfveenepae | SQL                  |
| rjqonwgagd | Informatica Generale |
| rjrbsovupf | SQL                  |
| rkpyhnweoz | Informatica Generale |

|             |                      |
|-------------|----------------------|
| rlbnjnuraa  | SQL                  |
| rqlhmwvvfr  | SQL                  |
| shwidpusut  | SQL                  |
| slfnwkouex  | SQL                  |
| soaxrpkmyx  | SQL                  |
| tahwadyybr  | SQL                  |
| tmalpukvph  | SQL                  |
| tmptxxfqlg  | SQL                  |
| twjbzfuhs   | SQL                  |
| uciubnhesv  | SQL                  |
| uctsvfnre   | Informatica Generale |
| ukteiawtcu  | SQL                  |
| uoujmoastk  | SQL                  |
| usloxptlnv  | SQL                  |
| vcetherudx  | SQL                  |
| vvesvpdxeu  | SQL                  |
| wbimqkxjpk  | SQL                  |
| wduzcrmbqr  | SQL                  |
| wgqmilmeeo  | SQL                  |
| wqoeigleak  | Informatica Generale |
| wyidhotyrc  | SQL                  |
| wyohxxjmbr  | SQL                  |
| xcvjgpwpgd  | SQL                  |
| xdfzuqqaxc  | SQL                  |
| xdpxeikdtg  | SQL                  |
| xxigdvhajv  | Informatica Generale |
| xymzioajbw  | SQL                  |
| yjakijsehd  | Informatica Generale |
| ynkrhlsqhx  | SQL                  |
| ypvazwngub  | SQL                  |
| yvbkrvodje  | SQL                  |
| zmivrwbcjja | SQL                  |
| zriwusbhcv  | SQL                  |
| zvsnjqiaqb  | SQL                  |

Tabella 12: Gruppi di argomento per le domande di logica

#### 4.5.2 Architettura della Rete testata

- ```
layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:120});
layer_defs.push({type:'fc', num_neurons: 2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 4, activation: 'tanh'});
```

```
layer_defs.push({type:'fc', num_neurons:10, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:120});

net = new convnetjs.Net();
net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});

• layer_defs = [];
layer_defs.push({type:'input', out_sx:1, out_sy:1, out_depth:120});
layer_defs.push({type:'fc', num_neurons: 2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons: 2, activation: 'tanh'});
layer_defs.push({type:'fc', num_neurons:10, activation: 'tanh'});
layer_defs.push({type:'regression', num_neurons:120});

net = new convnetjs.Net();
net.makeLayers(layer_defs);
trainer = new convnetjs.SGDTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, l2_decay:0.001});
```

**Prima architettura:** layers(2,4,10)

Il Reticolo della Conoscenza generato è il seguente:

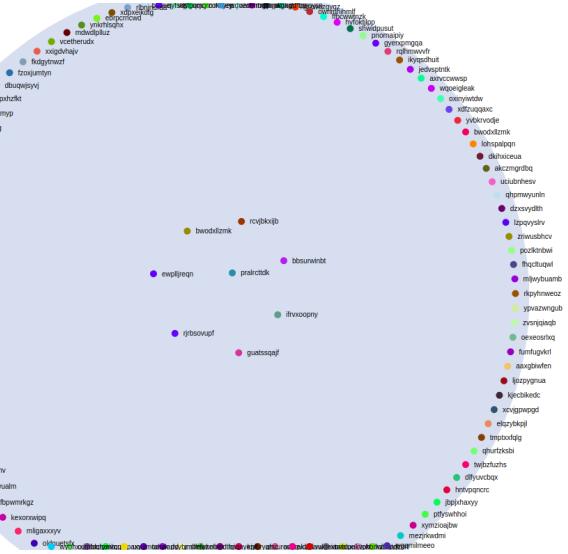


Figura 75: Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.



Figura 76: Reticolo della Conoscenza layers(2, 4, 10) - Forced Based.

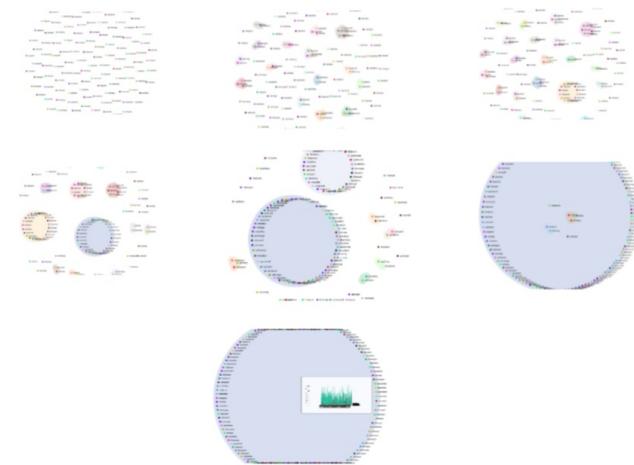


Figura 77: Sequenza di creazione del Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.

**Osservazioni dei risultati** Come si vede dalle immagini appena sopra, le ultime domande ad entrare a far parte del cluster sono:

Codice	Argomento
ifrvxoopny	SQL
pralrcttdk	Informatica Generale
rcvjbkxjbj	Informatica Generale

bwodxllzmk	SQL
ewplljreqn	Informatica Generale
rjrbsovupf	Informatica Generale
guatssqajf	SQL
bbsurwinbt	SQL

Tabella 13: Codice delle domande ad essere associate per ultime

Le domande di Informatica Generale vengono sufficientemente riconosciute dall’architettura come valori anomali.

### Seconda architettura: layers(2,2,10)

Il Reticolo della Conoscenza generato è il seguente:

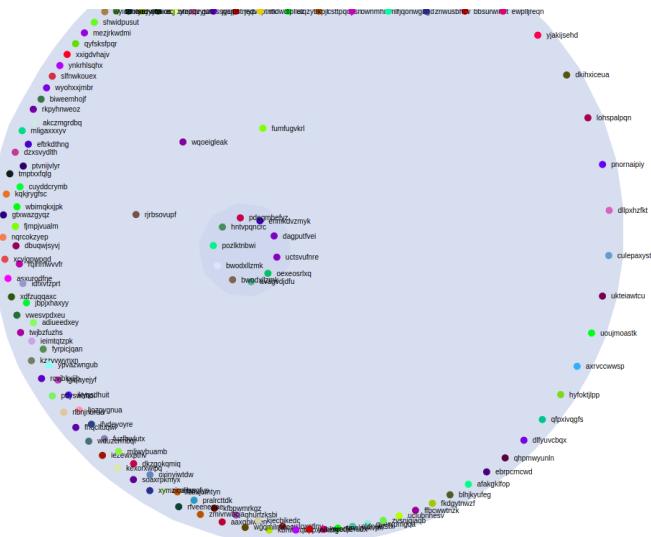


Figura 78: Reticolo della Conoscenza layers(2, 2, 10) - Cluster Based.

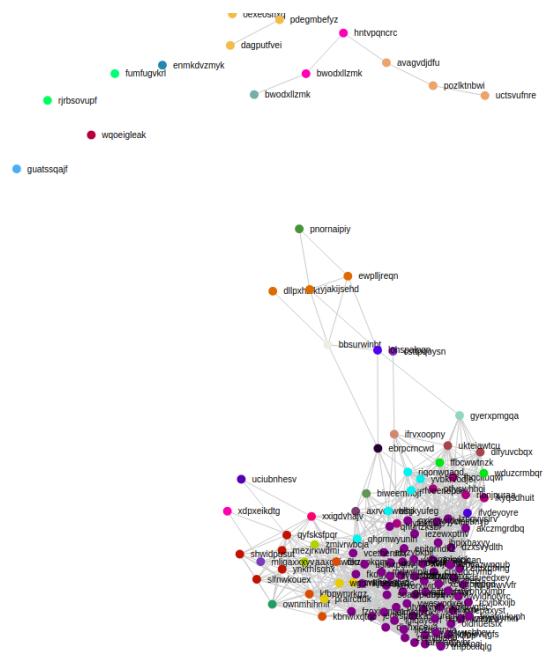


Figura 79: Reticolo della Conoscenza layers(2, 2, 10) - Forced Based.

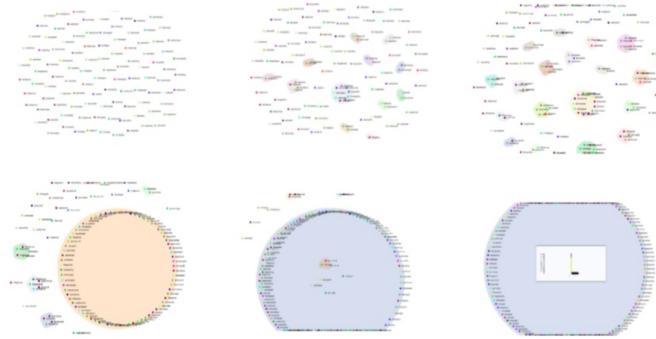


Figura 80: Sequenza di creazione del Reticolo della Conoscenza layers(2, 2, 10) - Cluster Based.

**Osservazioni dei risultati** Come si vede dalle immagini appena sopra le ultime domande ad entrare a far parte del cluster sono:

Codice	Argomento
enmkdvzmy	Informatica Generale
uctssvufnre	Informatica Generale
pdegmbefy	Informatica Generale

pozikthbw1	Informatica Generale
oexeosrxq	Informatica Generale
hntvpqncrc	Informatica Generale
dagputfvei	Informatica Generale
wqoeigleak	SQL
rjbsovupf	SQL
fumfugvkrl	SQL

Tabella 14: Codice delle domande ad essere associate per ultime

Le prime sei domande, contenute in tabella, appartengono ad un singolo cluster, nel quale ne risultano estranee solo le domande *bowdxllzmk* e *avagvdjdfu*, appartenenti alla categoria SQL.

Inoltre vengono associati in relazione stretta:

Codice	Argomento	Trattazione del quesito
oldnuetsfx fyrpicjcan	SQL	sargabile
usloxptlnv vcetherudx	SQL	selettività

Tabella 15: Codice delle domande associate

#### 4.5.3 Risultati di frequenza sulle domande SQL

Per maggiore completezza riporto il Reticolo della Conoscenza generato sulle risposte ai test SQL.



Figura 81: Reticolo della Conoscenza layers(2, 2, 10) con Frequenza - Cluster Based.

Il Reticolo generato dai dati di frequenza rende evidente la non capacità di un solo calcolo statistico di reagire ai valori anomali. Dalla figura sopra, si vede come le domande di Informatica Generale non vengono individuate e inserite all'interno del Reticolo solo quando si posiziona il threshold al massimo; ma anzi vengono individuate parzialmente, già nei primi stadi di formazione del reticolo, a gruppi in relazione a domande SQL, rientrando fin da subito all'interno del cluster principale.

Come pro della tecnica di frequenza è l'individuazione immediata di una domanda, *bwodxllzmk*, inserita due volte all'interno del set di dati.

cluter 2 (rosa)	cluster 3 (azzurro)
pozlktnbw	wqoeigleak
ewlpplljreqn	pralrcttdk
hntvpqncrc	pnomaipjy
oexeosrxq	fumfugvkr1
rcvjldkxib	rjrbsovupf
rkpyhnweo	

Tabella 16: Codice delle domande di logica associate a domande sql

Le domande che trattano la sargabilità di un predicato (*oldnuetsfx*, *fyrpicjgan*, *fuzfbwlutx*), non vengono messo in nessun modo in correlazione dal modello, come invece la Rete neurale effettua.

## 5 Configurazione dell'applicativo progettato e strumenti di sviluppo

### 5.1 Test d'unità

Per eseguire, con successo, i test d'unità è necessario procedere all'installazione, se non già presenti, dei moduli node\_js e delle componenti aggiuntive necessarie ad una corretta esecuzione degli script. Di seguito riporto le istruzioni per poter effettuare con successo la build dei test implementati<sup>11</sup>.

Prima di tutto l'utente deve aprire una *shell da terminale* e spostarsi all'interno delle sottocartelle rete\_neurale/rete neurale nella repository "AI-Reticolo\_della\_conoscenza". Per poter eseguire i test d'unità su qualsiasi sistema operativo, è necessarie procedere alle seguenti installazioni:

1. Installare node js da <https://nodejs.org/en/download/> per poter usare il comando npm;
2. Installare mocha: **npm install mocha**;
3. Installare il modulo babel core/register: **npm install --save-dev babel-core babel-preset-env**;
4. Installare presets per 2015: **npm install babel-cli babel-preset-es2015**;
5. Installare il modulo chai: **npm install chai**.

Per mandare in esecuzione i test digitare da terminale il comando **npm test**. Javascript è un linguaggio lato client, da browser. I test automatici sono stati sviluppati trasformando le singole unità in moduli node esportati. La sintassi exports poco si adatta all'utilizzo dei browser<sup>12</sup> per questo ho deciso di omettere le keywords exports come preambolo dei metodi soggetti a test; tale aggiunta è di competenza dell'utente nel momento precedente al test running.

---

<sup>11</sup>Se al esecuzione del comando da shall npm test viene generato un errore, seguire i seguenti passi di installazione

<sup>12</sup>Il supporto viene offerto da Safari 10.1, Chrome 61, Firefox 54, Edge 16

```

#Test1 - generator input
✓ generazione array mappato su grafo

#Test2 - generator input pure
✓ generazione array mappato su grafo

#Test3 - generator input probability
✓ generazione array spuro

#Test4
✓ array di colori uguali
✓ array di colori diversi

#Test5 - configurazione della rete andata a buon fine
✓ configurazione rete del database a 89 domande
✓ configurazione rete del database a 120 domande
✓ configurazione rete di prova

#Test6 - controllo valori inseriti nella form del differenziale di accoppiamento
✓ valore corretto entro i range stabiliti
✓ conversione a intero di un numero con virgola

#Test7 - stampa
✓ test del modulo di stampa

#Test8 - test matrice frequenza
✓ test frequenza con risposta positiva exists
✓ test frequenza con risposta positiva
✓ test frequenza con risposta negativa exists
✓ test frequenza con risposta negativa

33 passing (35ms)

```

Figura 82: Resoconto unit test con successo.

## 5.2 Sistema operativo e browser

Il progetto è stato sviluppato su sistema operativo Linux e testato sul sistema operativo Windows, messo a disposizione in ufficio. L'applicativo della Rete neurale è sviluppato ad hoc per browser Chrome in cui il caricamento di grandi moli di dati ha nettamente prestazioni migliori. Sconsiglio l'utilizzo dell'applicativo sviluppato su Microsoft Edge e Firefox , in quanto il caricamento del file di allenamento è molto lento <sup>13</sup>.

### 5.2.1 Specifiche di sviluppo

- Sistemi operativi di sviluppo:

---

<sup>13</sup>se il file ha dimensioni superiore a 1000 entry.

- Versione sistema operativo di sviluppo principale: Ubuntu 4.15.0-52.56-generic 4.15.18;
- Versione sistema operativo pc fisso aziendale: Windows 10 Enterprise, versione 1809;
- Browser di testing:
  - browser principale: Chrome versione 75.0.3770.100 a 64 bit e Versione 66.0.3359.170 a 64 bit;
  - browser (sconsigliato per l'uso della Rete del database): Microsoft Edge 44.17763.1.0, Microsoft EdgeHTML 18.17763;
  - browser (sconsigliato per l'uso della Rete del database): Firefox Quantum 67.0.2 a 64 bit.
- Strumenti di codifica:
  - Visual Studio Code 1.35.1.
- Strumenti di documentazione:
  - Texmaker 5.0.3 (compiled with Qt 5.9.5).

Il progetto è mantenuto all'interno di una piattaforma pubblica, con directory anch'essa pubblica. Sia il codice implementato che la documentazione sono soggetti a controllo di versione. L'indirizzo http dove la directory è posizionata, è il seguente:

[https://github.com/esignor/AI-Reticolo\\_della\\_conoscenza](https://github.com/esignor/AI-Reticolo_della_conoscenza)