

Resoconto Tecnico

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

Indice

1 Analisi dei dati di probabilità	4
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
2 Rete neurale	14
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 Natura delle domande di logica contenute nel database aziendale	23
2.2.3 Test e Documentazione	24
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
3 Principal Component Analysis	78
3.1 Metodologia applicata	78
3.2 Sviluppo	79
3.3 Risultati ottenuti	81
3.3.1 Calcolo della Principal Component Analysis	81
3.3.2 Calcolo degli autovettori ed individuazione di Summary	82
3.3.3 Calcolo degli autovettori	90
3.3.4 Calcolo della matrice di correlazione	90
3.4 Conclusione dell'analisi	93
4 Costruzione del Reticolo della Conoscenza	96
4.1 Descrizione del sistema	96
4.1.1 Configurazione	97
4.2 Creazione dei file CSV	98

4.3	Creazione del Reticolo della Conoscenza per sui dati di Prova	98
4.3.1	L'architettura della Rete	98
4.3.2	Reticolo della Conoscenza generato mediante previsioni sul set di dati che conto esclusivamente della conoscenza del candidato	99
4.3.3	Reticolo della Conoscenza generato mediante la frequenza calcolata sul set di dati che tiene conto esclusivamente della conoscenza del candidato	104
4.3.4	Osservazioni sul Reticolo generato dalle previsioni della rete	107
4.3.5	Reticolo della Conoscenza generato con set di dati che tiene conto della possibilità che un candidato ha di indovinare la risposta	108
4.4	Creazione del Reticolo della Conoscenza sui dati delle domande del database	109
4.4.1	Individuazione dei gruppi di argomenti	109
4.4.2	Architettura della Rete testata	112
4.4.3	Risultati di frequenza sulle domande di logica	122
4.5	Domande SQL	123
4.5.1	Individuazione dei gruppi di argomenti	123
4.5.2	Architettura della Rete testata	126
4.5.3	Risultati di frequenza sulle domande di logica	131
5	Configurazione dell'applicativo progettato e strumenti di sviluppo	133
5.1	Test d'unità	133
5.2	Sistema operativo e browser	134
5.2.1	Specifiche di sviluppo	134

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:

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

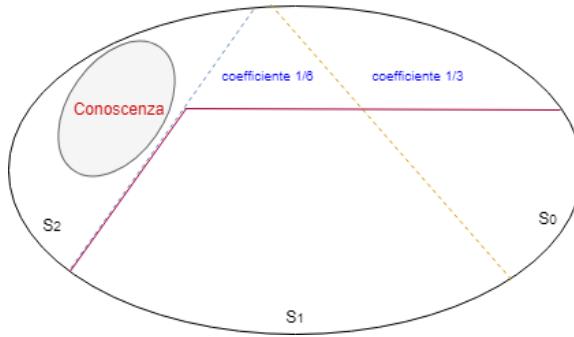


Figura 1: Rappresentazione insiemistica della probabilità di rispondere correttamente ad una domanda: $P(A)$

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 appoggia a rispondere ad una domanda può venire rappresentata in un piano.

Di seguito viene mostrata l'immagine di un modellino, rappresentativo di $P(A)$, realizzato durante l'analisi.

TODO: foto modello

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).

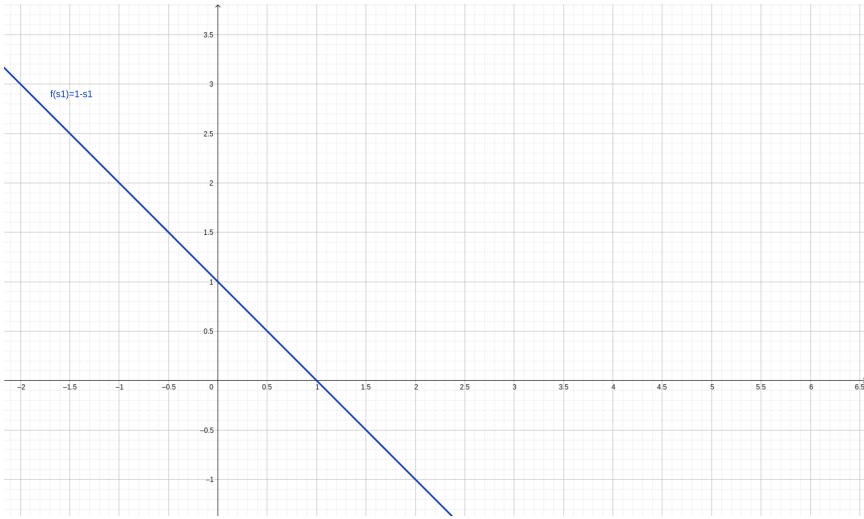


Figura 2: Rappresentazione della retta passante per $S_0 = 0$

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, ora, individuare il fascio di rette che tangenti il piano permettono di affermare se una specifica domanda è, in base alla sua frequenza, ha 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).

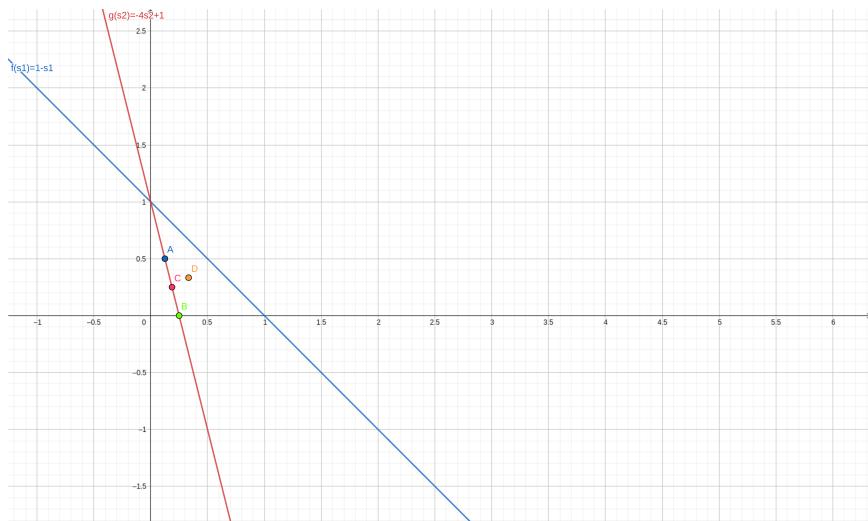


Figura 3: Rappresentazione di $P(A)$ per una frequenza 0.5 proiettata su assi $S_0 = 0$, S_1 e S_2 .

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 azzurra 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 né due risposte, perciò 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 - *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

```
Richiesta di previsione inviata alla rete ...
Richiesta 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... del 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 ...
Richiesta di previsione accettata
Il vettore [0,1,1,0,0,0] ha previsione calcolata di [-0.9703841414997074,-0.7824502014273779,0.34740891066895563,-0.8448991676224333,-0.5825080715009475,0.29369850449980084]
Rete Neurale in attesa ...
```

INSERIMENTO DATI DI ADDESTRAMENTO

Number of fields (max 9999): 6

Input1 Input2 Input3 Input4 Input5 Input6

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 ricevuti. Il problema che la rete mira ad analizzare è quello discusso nel precedente capitolo *Analisi dei dati di probabilità*

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 metodo, con il valore aggiunto, di generazione di un profilo randomico di 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 se nel periodo citato alle 89 domande di logica.

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 di tanti elementi quanti sono i neuroni di regressione dichiarati. Per la mia rete di prova è sufficiente dichiarare 6 neuroni in regressione perché 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.

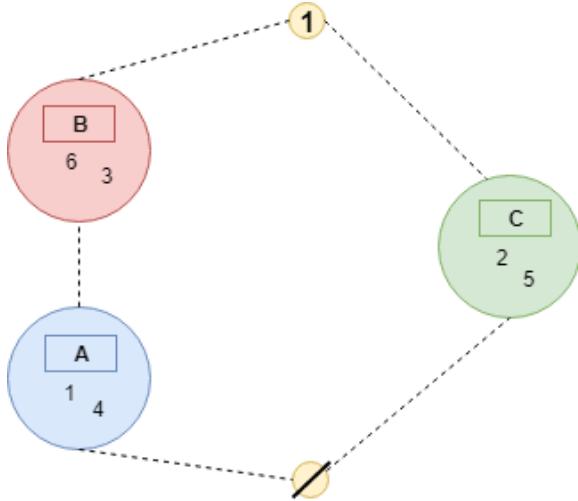


Figura 6: Grafo rappresentante le relazioni esistenti tra il set di domande di prova.

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 correndo il rischio di avere una rete che apprende troppo e come effetto negativo "veda" addirittura cose che non esistono. A prova di ciò sono i risultati non conformi alle attese. Dunque mi fermo qui con il test di tale architettura e riducendone il numero di neuroni presenti in ciascun layers e/o il numero di layers presenti.

Le nuove architetture su cui ho effettuato i test sono esposte nei paragrafi seguenti.

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 più di non saper saper rispondere alle 3 e 6 (come imposto dal vettore previsione).

Training set con generazione del profilo di un candidato e calcolo delle probabilità di risposta 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 che sul grafo della conoscenza.

- 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 che sul grafo della conoscenza.

- 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 che sul grafo della conoscenza.

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 emerso di discordate dal 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; causa di un numero eccessivo di neuroni dati alla rete per l'apprendimento rispetto al training set svolto. La situazione è generatrice di overfitting e non attendibilità dei dati raccolti. L'architettura a 1 e 2 neuroni 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 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 è 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).

```
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:3});

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

Inserito vettore di risposta: [1,-1,1,1,-1,1]
Inserito vettore di risposta: [1,1,1,1,1,1]
Inserito vettore di risposta: [1,1,1,1,0,1]
Inserito vettore di risposta: [1,-1,1,1,0,-1]
Inserito vettore di risposta: [-1,1,1,1,0,1]
Inserito vettore di risposta: [-1,1,1,1,1,0]
Inserito vettore di risposta: [-1,1,1,1,1,-1]
Inserito vettore di risposta: [1,1,1,1,-1,1]
Inserito vettore di risposta: [0,1,1,0,1,1]
Inserito vettore di risposta: [0,1,1,0,1,-1]
Inserito vettore di risposta: [1,-1,1,1,1,1]
Inserito vettore di risposta: [1,0,1,1,0,1]
Inserito vettore di risposta: [0,1,1,1,0,1]
Inserito vettore di risposta: [0,1,-1,1,1,1]
Inserito vettore di risposta: [1,-1,1,1,1,1]
Inserito vettore di risposta: [1,0,1,1,1,1]
Inserito vettore di risposta: [1,0,1,1,1,-1]
Inserita richiesta di prevedere...
Richiesta di prevedere inoltrata alla rete ...
Il vettore [0,0,0,0,0] ha previsione calcolata di [0.45968601518151, -0.4194727901763507, -0.3865221544720565, 0.44336712317213167, 0.40599951266191436, -0.333082413748274]
Rete neurale in attesa ...
```

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

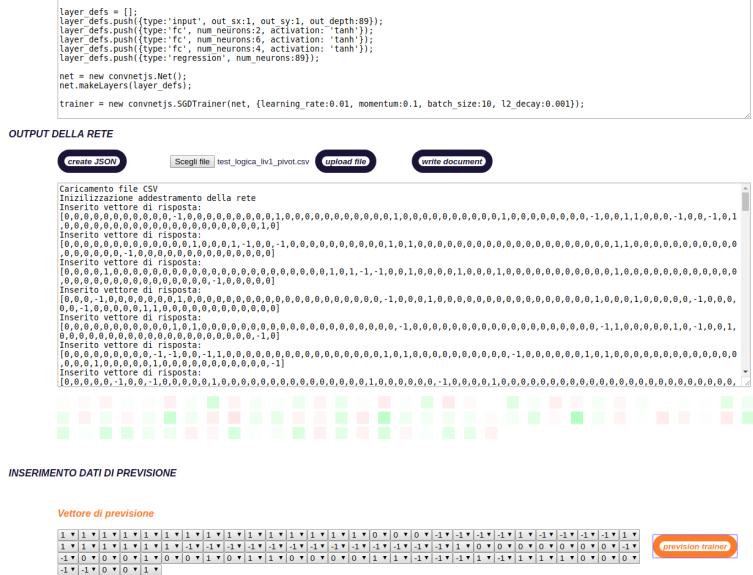


Figura 8: Rete neurale del database aziendale.

Una prima architettura su cui ho deciso di analizzare i risultati della rete del database a 89 domande di logica, basandomi anche sul quanto appreso dalla Rete neurale di prova e dal 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 Natura delle domande di logica contenute nel database aziendale

Analizzando il training set dei vettori ho riscontrato tali correlazioni:

- Solo una piccola parte delle domande di logica, presenti nel database, vengono svolte durante un colloquio con un candidato. In media una decina su 89 possibili.
- Dalla rete sembra che le domande abbiano qualche correlazione, tuttavia la configurazione attuale ne rende difficoltosa l'individuazione.

2.2.3 Test e Documentazione

Durante la settimana dal 10/06 al 18/06 ho effettuato quanto definito all'interno del Piano di Lavoro come "Test e Documentazione".

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
 $[0.8521066399598267, 0.898137375081856, 0.9993098151218291, 0.792190337086403,$
 $0.811145866789799, 0.9514731560722426]$

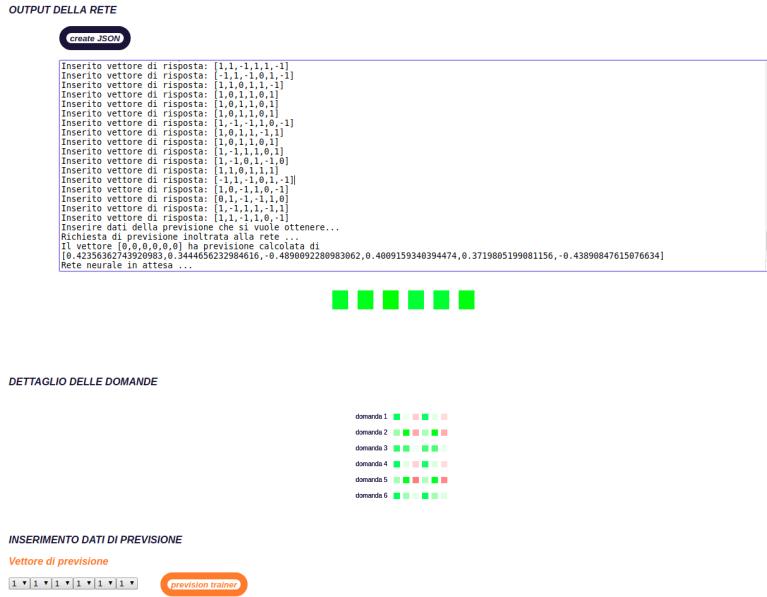


Figura 9: Risultato della rete di prova a seguito di un vettore di previsione [1, 1, 1, 1, 1, 1] in input.

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 comunque 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 basso di venire risposto 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 ¹.

Tuttavia non sono, durante questo periodo, 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;

¹valutando i cluster generati

- la creazione stessa del set di input, parzialmente randomica.

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.SGDTtrainer(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.

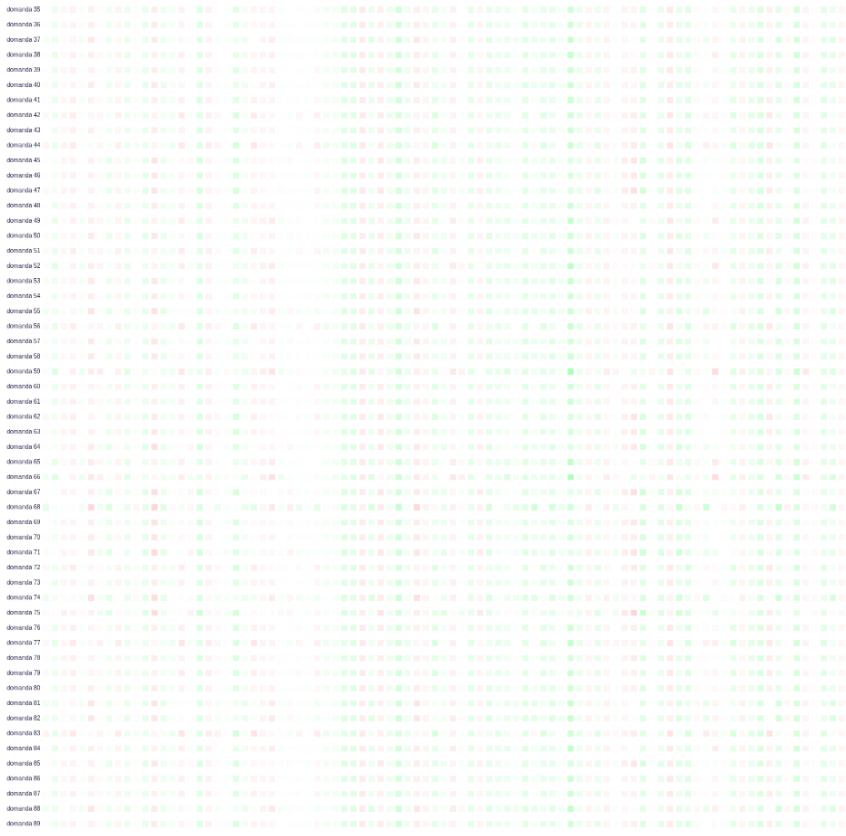


Figura 12: Risultato della rete del database a seguito di un vettore di previsione $[1, 1, 1, 1, 1, 1]$ in input.

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 in 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.

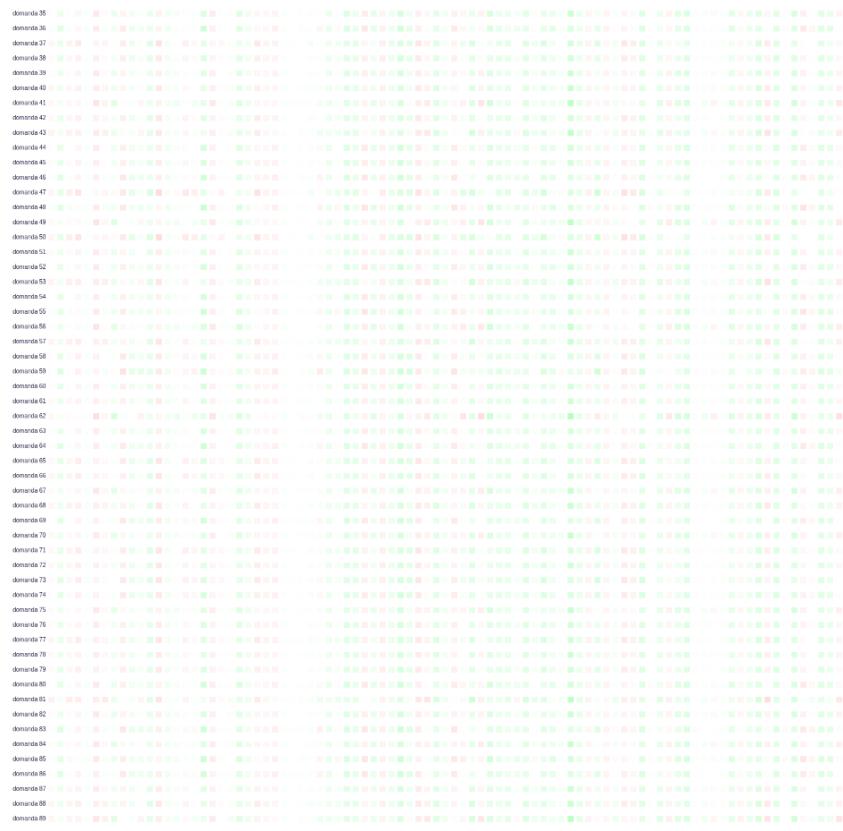


Figura 14: Risultato della rete del database a seguito di un vettore di previsione $[-1, -1, -1, -1, -1, -1]$ in input.

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.

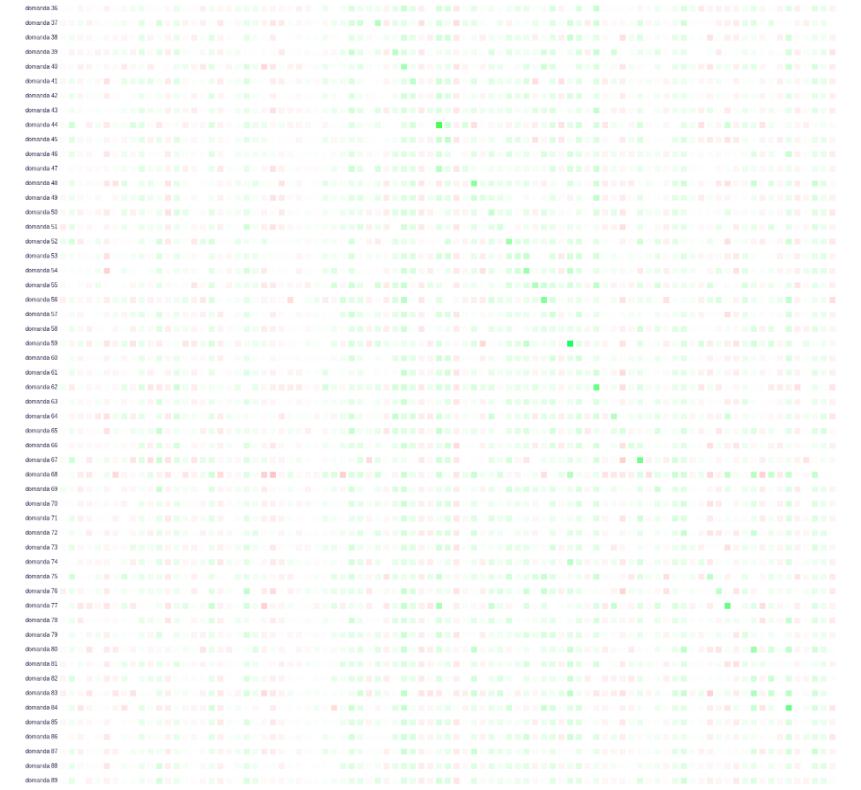


Figura 16: Risultato della rete del database a seguito di un vettore di previsione $[1, 1, 1, 1, 1, 1]$ in input.

Dagli screen della rete riportati sopra appare come la situazione appare meno lineare del caso analizzato precedentemente. Le domande non vengono separate per linee rete; 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 precedentemente.

Architetture sottoposte ad analisi:

```

• 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 in merito all'architettura di 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 di 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. Se si desidera modificare il numero di domande analizzate dalla rete è d'obbligo modificare il numero di neuroni espressi nella regressione, di previsione, di configurazione e index della rete;

- 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 dalla numerazione appena descritta sopra seguita da una stringa testuale che rappresenta la funzionalità che si vuole modellare;
- **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 è 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

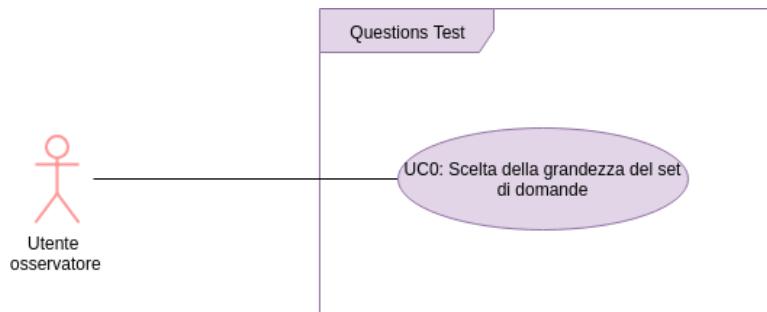


Figura 17: UC0_g: Operazioni utente - Visione generale della pagina di accesso alla Rete - Questions Test

- **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

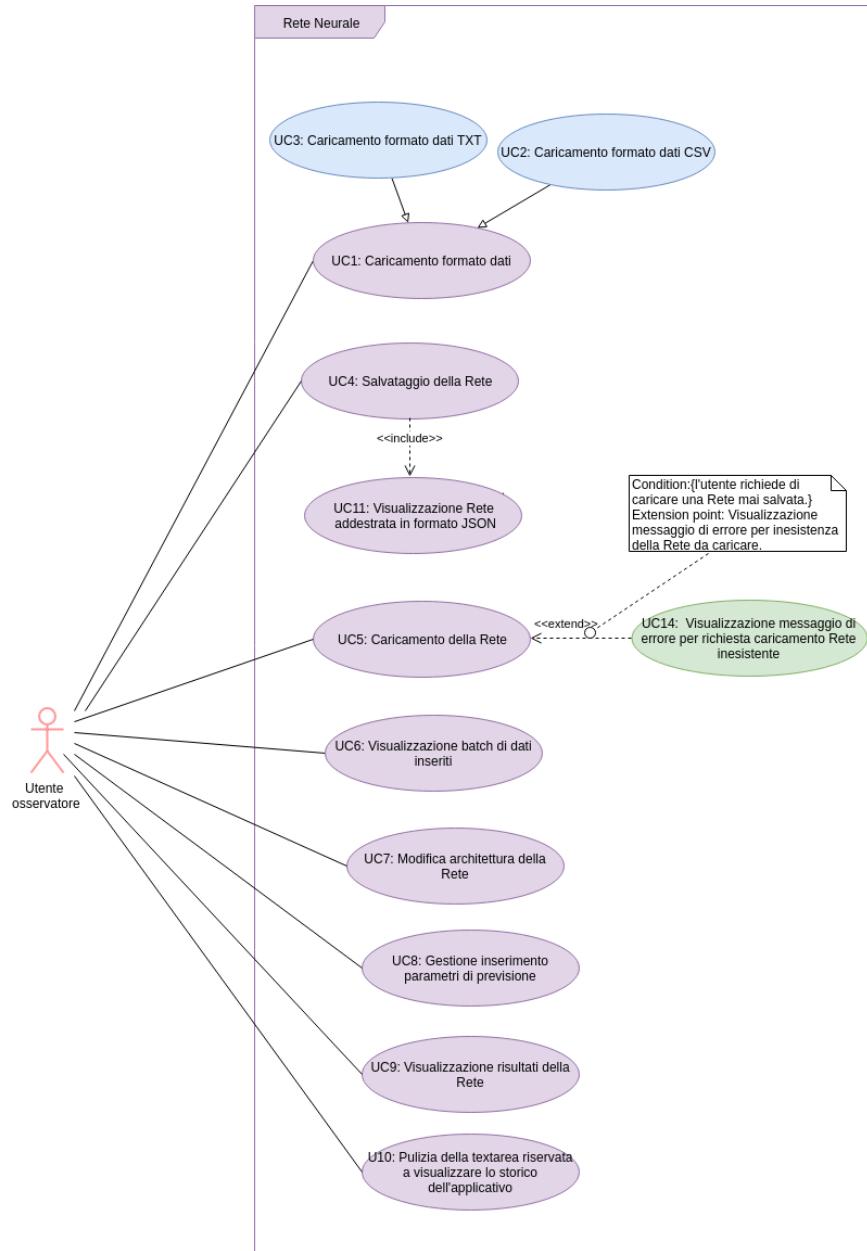


Figura 18: UC1-g: Operazioni utente - Visione generale interfacciamento con la Rete neurale

- **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;

²modello 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
Funzionale	10	12	6	28
di Vocolo	1	3	1	5
di Qualità	1	0	0	1
Totale	12	16	7	35

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 ≥ 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 inserita 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.

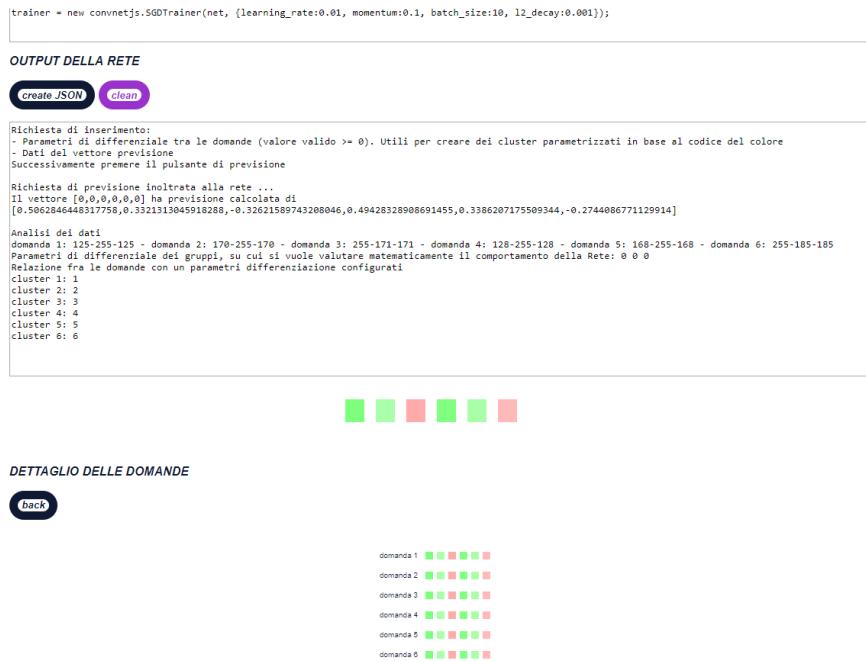


Figura 20: Interfaccia finale utente della Rete neurale di prova - Visione canvas generale e di dettaglio.

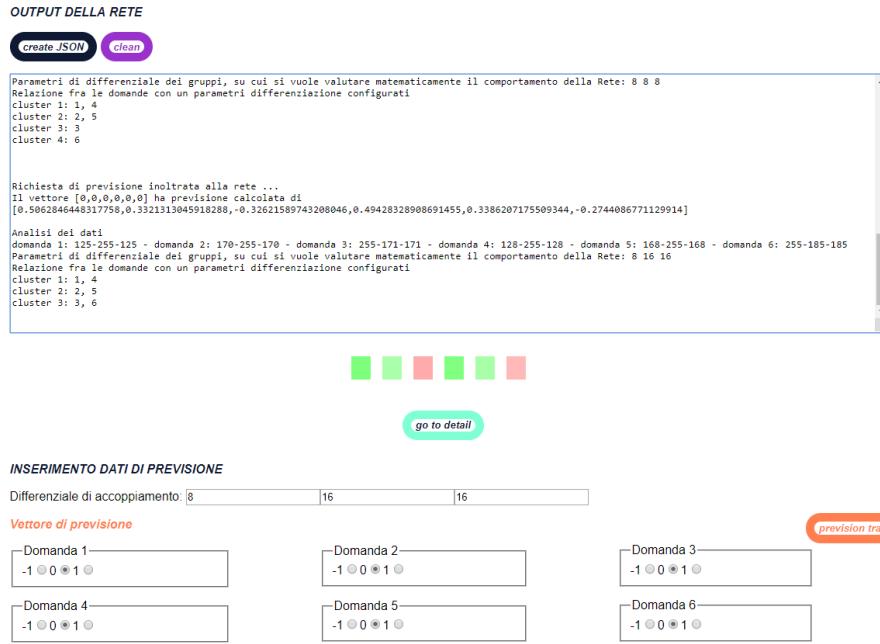


Figura 21: Interfaccia finale utente della Rete neurale di prova - Visione inserimento parametri di previsione.

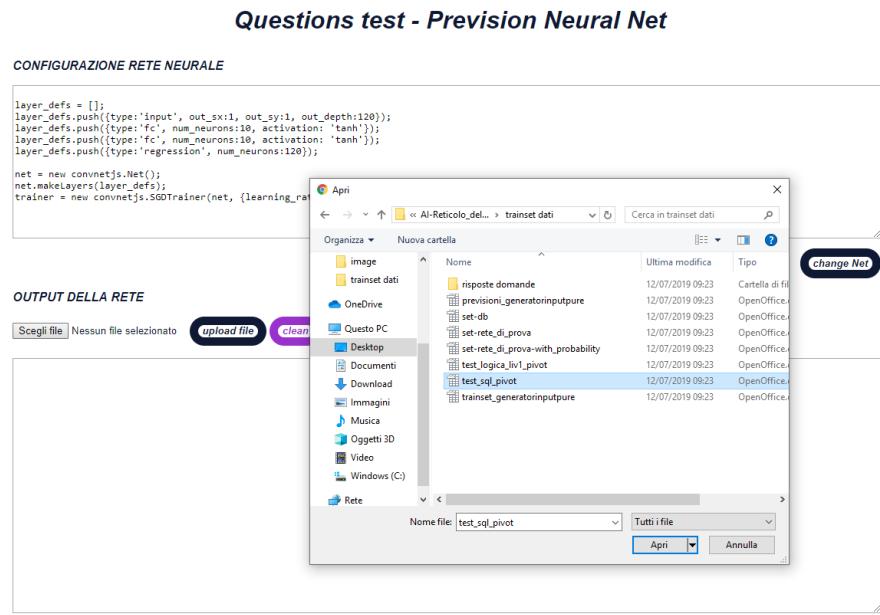


Figura 22: Interfaccia finale utente della Rete neurale del database - Visione caricamento file dati.

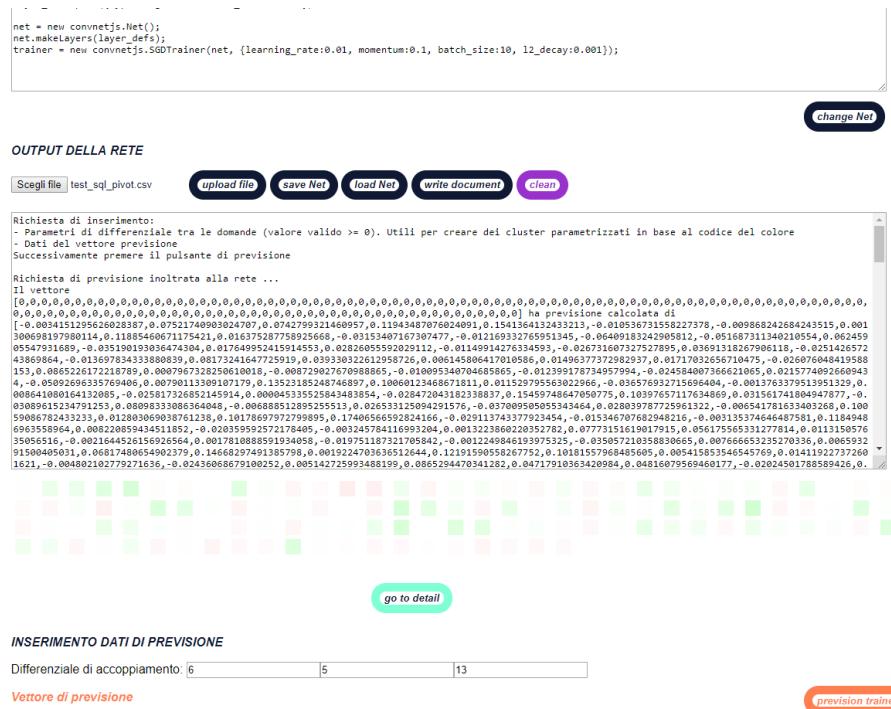


Figura 23: Interfaccia finale utente della Rete neurale del database - Visione box con storico delle operazioni e canvas generale.

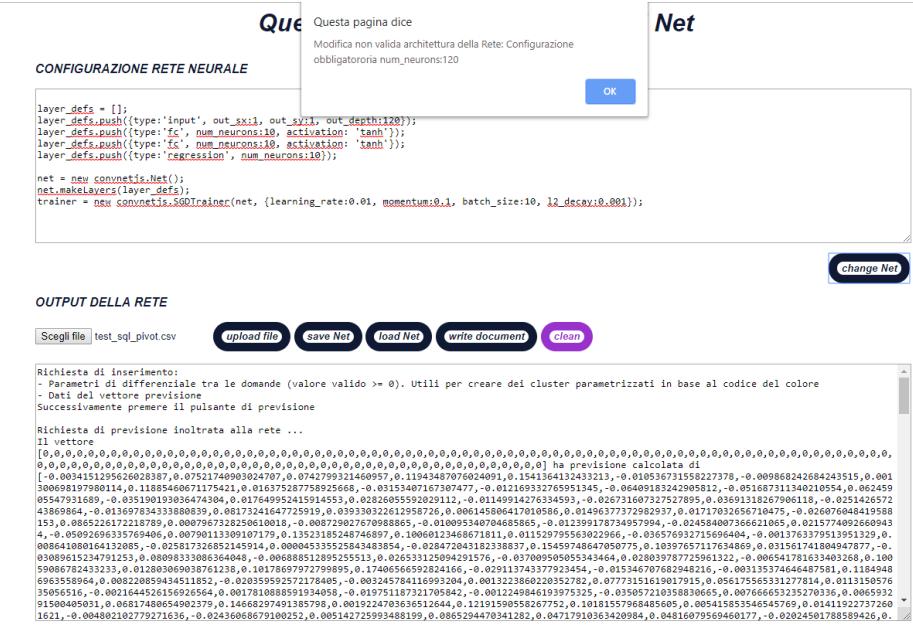


Figura 24: Interfaccia finale utente della Rete neurale del database - Visione errore rilevato durante la configurazione dell'architettura.

3 Principal Component Analysis

La **Principal Component Analysis** (acronimo PCA) è una tecnica impiegata nell'ambito della statistica multivariata³ 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;

³parte 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 perché 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

I dati risultati dall'elaborazione dei dati inerenti le domande nel database gli ho riportati, nelle sezioni seguenti, per motivi di spazio e comprensione, in modo parziale. Ogni metodo utilizzato viene di seguito descritto nel dettaglio.

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 0.00111520653 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 25: Visualizzazione del calcolo della PCA tramite prcomop per il set di prova.

```
> prcomp(df_numeric, scale = FALSE)
Standard deviations (1, ..., p=9):
[1] 0.58043495 0.55808746 0.54971833 0.53126069 0.52059133 0.51577310 0.49349655 0.48596238 0.47999862
[2] 0.42079249 0.41553049 0.41868652 0.40526268 0.40119893 0.39676359 0.38988115 0.38394838 0.38248451
[3] 0.34271823 0.33861509 0.33532453 0.33275471 0.32925297 0.32449961 0.31762233 0.31405458 0.30962208
[4] 0.28077829 0.27888496 0.27585038 0.27422473 0.27139698 0.27019256 0.26679640 0.26427259 0.26314990
[5] 0.24187129 0.23791610 0.23498756 0.23280141 0.22959808 0.22384654 0.21341210 0.19833321 0.180912521
[6] 0.14016744 0.13282731 0.10140180 0.08865511 0.02703589 0.02534545

Rotation (n x k) = (89 x 9):
PC1          PC2          PC3          PC4          PC5          PC6          PC7          PC8          PC9          PC10         PC11         PC12         PC13         PC14
V1 2.436349e-03 0.005200166 -4.597495e-03 0.084770595 -6.267518e-04 0.001175541 -0.000660271 0.001555997 -1.338486e-03 0.000166171 -0.00320794 0.013871066 -0.012466570 0.01606655
V2 2.024064e-02 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000 0.000117000
V3 2.415304e-03 -0.007382031 1.237214e-02 -0.0130787491 3.664898e-03 -0.017104077 0.813993561 -0.0076250979 -1.464139e-02 -0.0264389810 -0.015997338 0.0123222475 -0.0312085902 -0.007941486
V4 -6.730258e-03 -0.015745591 2.178088e-02 -0.0117814649 4.127785e-02 0.0022788841 -0.008823932 7.185564e-03 0.0081277662 -0.0081277662 -0.0081277662 -0.0081277662 -0.0081277662 -0.0081277662 -0.0081277662
V5 1.514325e-02 -0.005739754 6.121892e-03 0.00645128218 5.433210e-03 0.0005548622 -0.018518277 -0.0083915065 5.902556e-03 0.0027513825 0.009012403 0.0399523132 0.00801994258 -0.007935540
V6 3.397826e-02 0.0335866990 1.197013e-02 -0.0933261778 2.604523e-05 0.0141769088 0.0067276127 -0.0041578833 -2.553817e-02 -0.0384170573 0.0118988680 -0.0188077858 0.0477093378 0.022818524
V7 1.314959e-02 0.00016794227 -3.653464e-02 0.0332199318 4.94064e-02 0.0253328818 -0.0212349059 7.94952e-03 0.00016794227 -0.01822349059 -0.01822349059 -0.01822349059 0.00016794227 0.00016794227 0.00016794227
V8 1.273740e-02 0.00016794227 -3.653464e-02 0.0332199318 4.94064e-02 0.0253328818 -0.0212349059 7.94952e-03 0.00016794227 -0.01822349059 -0.01822349059 -0.01822349059 0.00016794227 0.00016794227 0.00016794227
V9 4.916774e-02 0.0313620891 -0.155063e-02 0.163807886 2.144642e-01 0.0159534515 0.0340172249 -0.0240780166 -7.43864752e-02 0.10162547355 -0.1932998373 0.1526616739 0.044153589 0.07809444
V10 1.203660e-02 0.0063310939 -1.25139e-02 -0.0241742814 -2.706798e-02 -0.0193394845 -0.0222724247 0.0529256183 -0.0389252821 -0.029166157
V11 2.707129e-02 0.1619149972 -1.111640e-02 0.1515748131 -0.052191e-02 0.0595648752 -0.0584362752 -0.1255892e-02 0.3334747699 0.1520353897 -0.0680831778 -0.0499464942 -0.326270930
V12 2.441796e-04 -0.0088153860 -0.168016e-02 -0.0395959227 0.0533941e-03 0.0165549106 0.0464409945 0.0549463204 0.04989919258 -0.00757255
V13 2.441796e-04 -0.0088153860 -0.168016e-02 -0.0395959227 0.0533941e-03 0.0165549106 0.0464409945 0.0549463204 0.04989919258 -0.00757255
```

Figura 26: 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 27: Visualizzazione del calcolo del metodo summary ed individuazione degli autovalori per il set di prova.

```
> summary(res.pca)
Importance of components:
          Corp.1    Corp.2    Corp.3    Corp.4    Corp.5    Corp.6    Corp.7    Corp.8    Corp.9    Corp.10   Corp.11   Corp.12   Corp.13   Corp.14   Corp.15   Corp.16
Standard deviation 1.43136144 1.30933969 1.27234891 1.25540778 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 0.016168611 0.015643872 0.01534701 0.01506496 0.01504464 0.01484788
Cumulative Proportion 0.02103668 0.04231675 0.06695249 0.07824758 0.09578028 0.11269334 0.12948710 0.14617321 0.16261151 0.17861659 0.1942427 0.2249932 0.2406328 0.25510814 0.26995692
          Corp.17   Corp.18   Corp.19   Corp.20   Corp.21   Corp.22   Corp.23   Corp.24   Corp.25   Corp.26   Corp.27   Corp.28   Corp.29   Corp.30   Corp.31   Corp.32
Standard deviation 1.10319196 1.12886362 1.12044621 1.11484745 1.10424461 1.09340381 1.08375051 1.07395459 1.06356189 0.99140845 0.97984747 1.07000985 1.06700985 1.06400985 1.06100985
Proportion of Variance 0.018452855 0.018033055 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421 0.018424421
Cumulative Proportion 0.028452855 0.29865910 0.31307331 0.32704943 0.34081338 0.35457041 0.36813229 0.38152687 0.39473653 0.40783165 0.42067653 0.43376981 0.44647318 0.45939312 0.47158093 0.48393076
          Corp.33   Corp.34   Corp.35   Corp.36   Corp.37   Corp.38   Corp.39   Corp.40   Corp.41   Corp.42   Corp.43   Corp.44   Corp.45   Corp.46   Corp.47   Corp.48
Standard deviation 1.04333725 1.04116609 1.02871449 1.02479858 1.02307959 1.01727836 1.01559804 1.00733966 1.00662121 0.99742429 0.99650091 0.99570404 0.98376459 0.98285548 0.97349201 0.97125294
Proportion of Variance 0.012124076 0.012118987 0.01199085 0.0118962 0.01177605 0.01163706 0.01159812 0.01149866 0.01139439 0.01118753 0.0111643 0.01114863 0.01108282 0.010945672 0.010845971
Cumulative Proportion 0.034453392 0.056571223 0.078759175 0.099976378 0.120197202 0.142322338 0.162338389 0.182346671 0.202354671 0.222352208 0.242359127 0.26236647 0.282373121
          Corp.49   Corp.50   Corp.51   Corp.52   Corp.53   Corp.54   Corp.55   Corp.56   Corp.57   Corp.58   Corp.59   Corp.60   Corp.61   Corp.62   Corp.63
Standard deviation 0.96334887 0.95952978 0.95755893 0.95339737 0.941994693 0.9325598 0.932384719 0.931159665 0.92139027 0.918259719 0.907151736 0.901228296 0.894532247 0.893069712
Proportion of Variance 0.018047919 0.018035323 0.018131874 0.01822132 0.0189976378 0.018999143 0.019977054 0.0189749859 0.019954468 0.01898481683 0.019944323 0.0189253773 0.010913317 0.008998101 0.00899704
Cumulative Proportion 0.67685352 0.68720863 0.69751937 0.707744668 0.717171968 0.727616203 0.737390257 0.747146189 0.75668946 0.776166203 0.77569946 0.7848346 0.79499419 0.803963523
          Corp.75   Corp.76   Corp.77   Corp.78   Corp.79   Corp.80   Corp.81   Corp.82   Corp.83   Corp.84   Corp.85   Corp.86   Corp.87   Corp.88   Corp.89
Standard deviation 0.887804612 0.8800884930 0.877852651 0.871862307 0.865451738 0.859708007 0.856741253 0.847836174 0.846571589 0.837088957 0.829688903 0.82365230 0.817133935 0.812411088 0.808770848
Proportion of Variance 0.009886329 0.008707979 0.008649888 0.008547944 0.0089421851 0.008311147 0.008233881 0.0080923310 0.008052626 0.0079749849 0.007828646 0.007512879 0.007421824 0.007355447
Cumulative Proportion 0.820828816 0.829536613 0.8318186499 0.846734442 0.8515156293 0.863467440 0.871712324 0.879804515 0.887867164 0.895746728 0.903487576 0.911116222 0.918628301 0.926050125 0.933405572
          Corp.80   Corp.81   Corp.82   Corp.83   Corp.84   Corp.85   Corp.86   Corp.87   Corp.88   Corp.89
Standard deviation 0.79836916 0.796146694 0.782386719 0.772551946 0.759737073 0.755173864 0.745553834 0.728101546 0.717568684 0.699402168 0.663042160 0.624248461
Proportion of Variance 0.007167148 0.007127629 0.006883383 0.006711419 0.0064969511 0.006421863 0.006217062 0.005963326 0.005798608 0.005402160 0.0052438461
Cumulative Proportion 0.94057305 0.947786081 0.954584963 0.961295482 0.967786093 0.974198956 0.980416018 0.986377344 0.992187413 0.997569319 1.000000000
```

Figura 28: 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 29: 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.

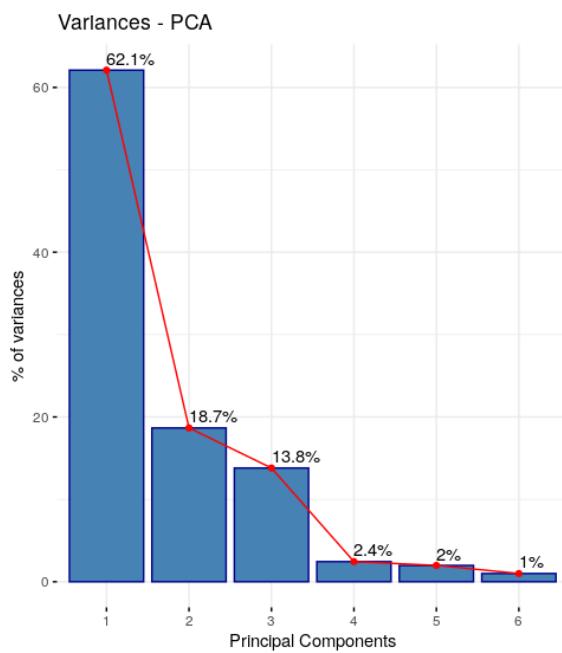


Figura 30: Plot della rappresentazione grafica della varianza sui PC del set di prova.

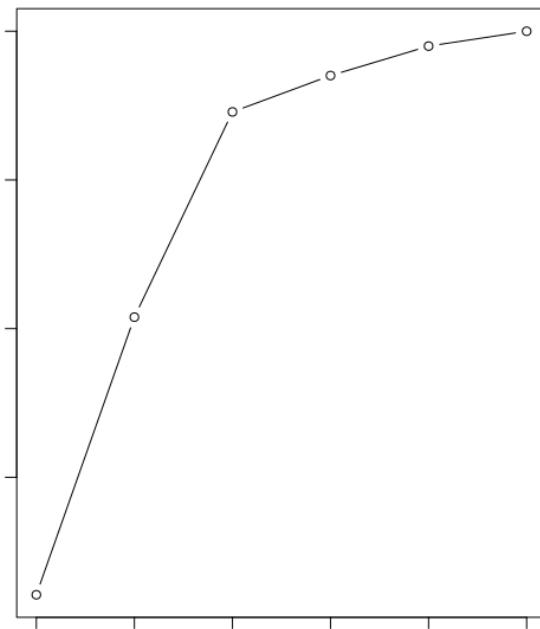


Figura 31: Plot della rappresentazione grafica della varianza sui PC del set di prova.

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

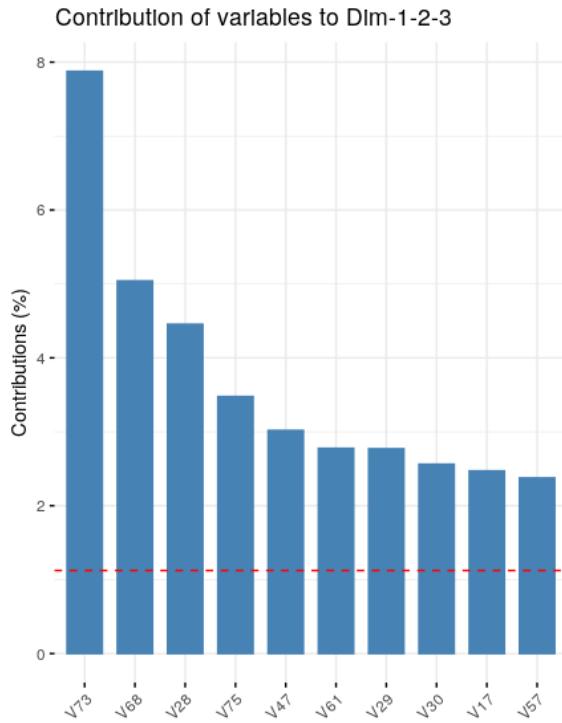


Figura 32: Rappresentazione grafica di come la varianza si distribuisce sulle PC individuate dal modello sul set di prova.

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.

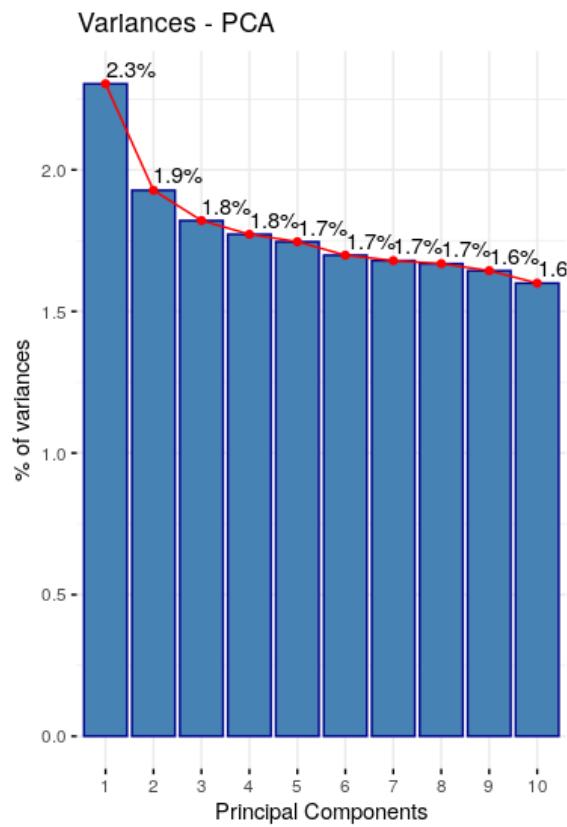


Figura 33: Plot della rappresentazione grafica della varianza dei primi dieci PC del set delle domande nel database.

Il plot 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.

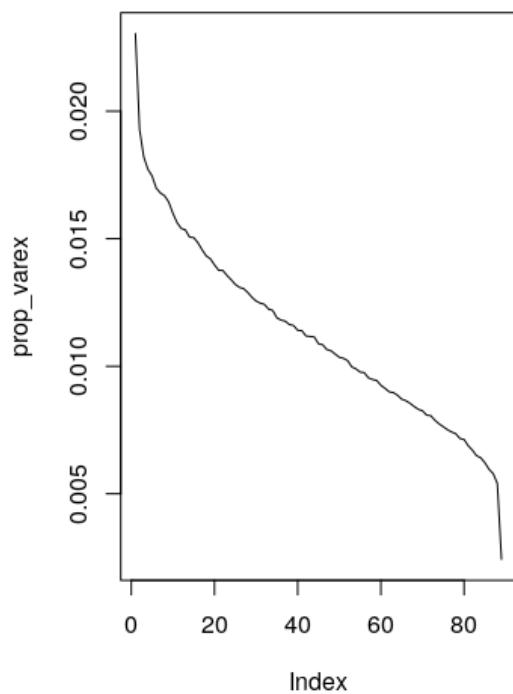


Figura 34: Plot della rappresentazione grafica della varianza di tutti i PC del set delle domande nel database.

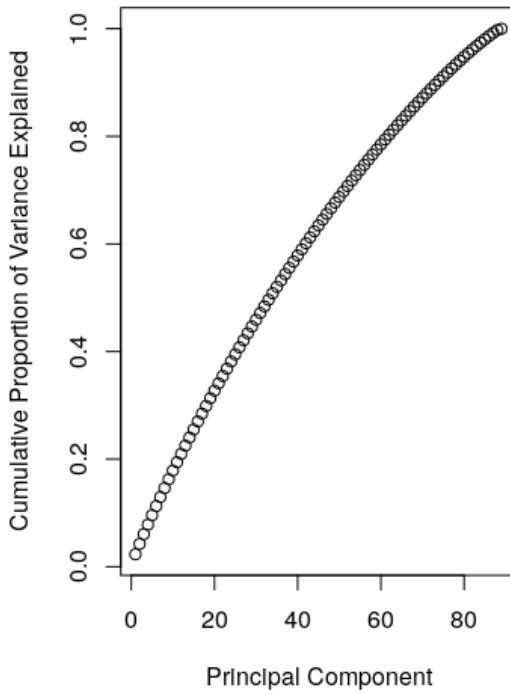


Figura 35: Plot della rappresentazione grafica della varianza di tutti i PC del set delle domande nel database.

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 CP 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 CP. 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 CP 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).

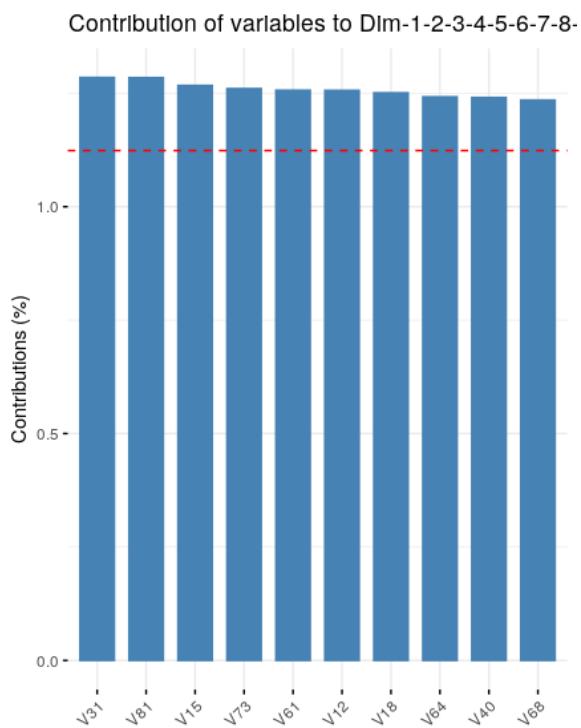


Figura 36: Rappresentazione grafica di come la varianza si distribuisce sulle PC individuate dal modello sul set delle domande nel database.

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 37: 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 nel database; 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  1.0000000 -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 38: Autovettori del set di test.

Figura 39: Autovettori del set delle domande nel database.

I CSV generati di correlazione sono i seguenti

	A
1	V1" "V2" "V3" "V4" "V5" "V6
2	"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 40: CSV generato a partire dalla matrice correlazione del set della rete di prova.

Figura 41: CSV generato a partire dalla matrice correlazione del set delle
distanze dei dati.

- le CSV gli ho generati con i seguenti passaggi:

 - Calcolando 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 analisi dei dati delle domande di logica, presenti nel database, 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, contenute nel database aziendale 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 pot 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, nel database, 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.

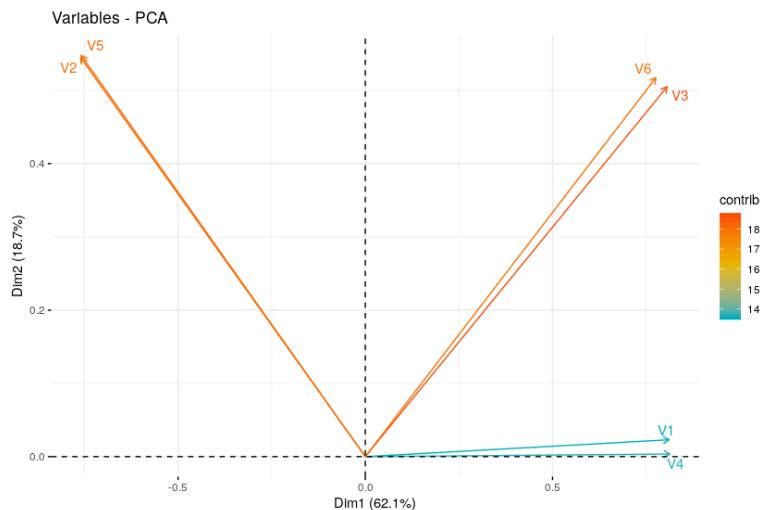


Figura 42: Rappresentazione per mezzo di plot di come le variabili si presentano nelle due componenti principali con il calcolo della PCA - utilizzo del set di prova puro sul grafo della conoscenza.

```
> 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 43: Le domande in correlazione con la variabili principali in ordine decrescente sui dati di prova.

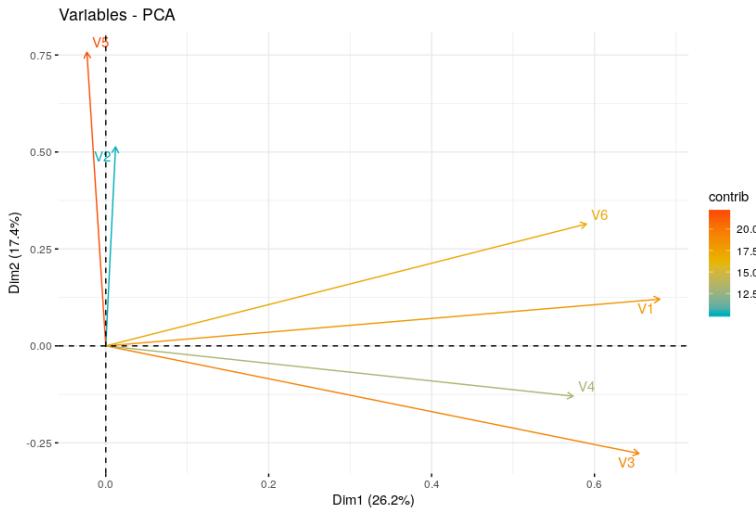


Figura 44: Rappresentazione per mezzo di plot di come le variabili si presentano nelle due componenti principali con il calcolo della PCA - utilizzo del set di prova spurio con la probabilità di indovinare.

```
> 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 45: 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 costruito, 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.

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 standard⁴.

⁴Vettore previsione tutto a zero

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

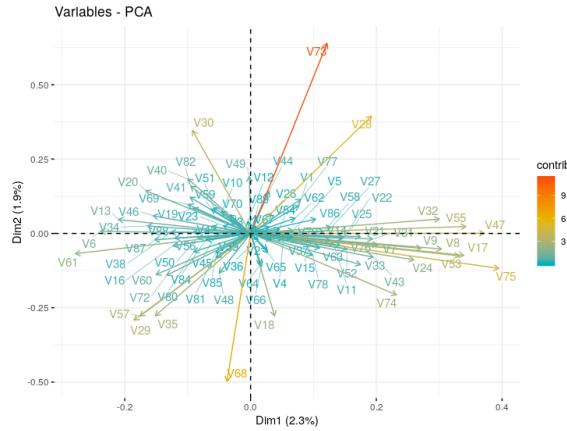


Figura 46: Le domande in correlazione con le variabili principali in ordine decrescente sui dati di prova soggetti alla probabilità di indovinare.

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 import 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 gli 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 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 è la seguente:

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

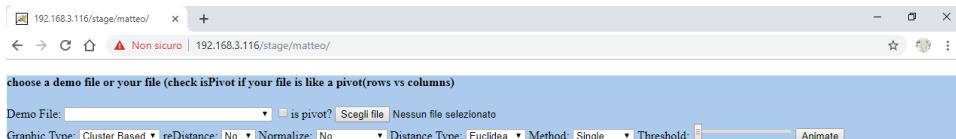


Figura 47: Configurazione usata nel sistema per generare il Reticolo della Conoscenza.

4.2 Creazione dei file CSV

Come già accentato all'interno della sezione §Descrizione del sistema 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 neurale oggetto di studio (di prova o del database), di calcolare:

1. Le previsioni ottenibili da un vettore di previsione settato a 1 o -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 ⁵.

Ho fatto in modo che il vettore delle differenze venga stampato su console del browser ⁶, 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;

⁵Vettore tutto a zero

⁶unica 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 modello⁷ 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:

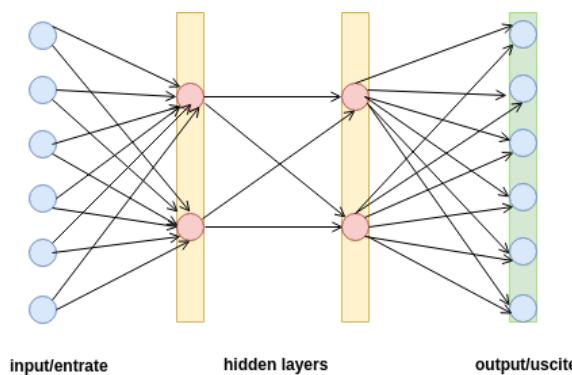


Figura 48: Architettura delle Rete di prova: 6 ingressi in una rete a 3 strati con 2 stati nascosti e uscite.

A	B	C	D	E	F	G	H	I	J	K	L	M
1	2	3	4	5	6	7	8	9	10	11	12	13
2	3.171737411696005329601349858143	3.0718504387564600753891112308602959	3.12366620179563979	3.070777313135399	1.129314762326576	0.281924486022983	0.00046039394708	0.1392350254466009	0.281727298323205	0.36156486009116974		
3	2.033244695752652219840442017289847	0.16314499371485910	0.220055835406341729	0.48612688776466211	0.1557138786077339	0.36833974152365005	0.625905456612903	0.101999886170880	0.370988604411429	0.5702448188770408	0.36946459200033014	
4	3.0159886032661289	0.111430322447592	0.683521671569856	0.1511047120835172	0.12372266756852643	0.109517191689329	0.105033781425388	0.039365624297840	0.140766413661771	0.4790412057547656		
5	4.023405155777997470	0.20991016847346	0.468383169399651	0.23102086435170	0.25819239041250718	0.43501781616958	0.196149522328257	0.23312997581812	0.3677864164962118	0.139403571014482	0.2317432649958	0.3471596810572275
6	5.00871117987591130	0.488929446693123	0.396155675269853	0.476223037076929	0.50455271645051	0.479655811971496	0.197774748128393	0.383797139990685	0.3444112501920	0.1256521521054321	0.598690302696974	0.27122593023303
7	0.12592225909135909	0.024492988708170	0.20977415078975	0.218796413991359	0.0413900148914899	0.7464331009211	0.153872481086939	0.046879324171877	0.144918550492490	0.0984007569571200	0.0423827150703	

Figura 49: file CSV generato per la creazione del Reticolo della Conoscenza sui dati di prova.

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

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

⁷inteso come architettura.



Figura 50: Reticolo della Conoscenza per i dati di prova - Cluster based.

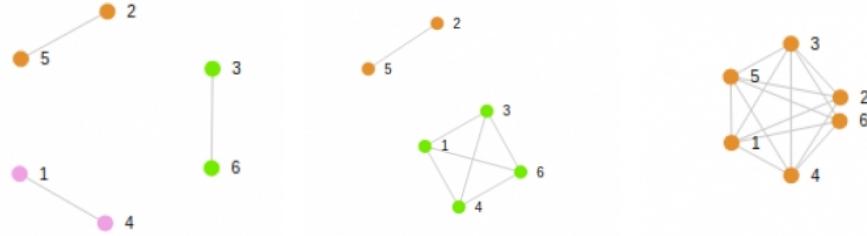


Figura 51: Reticolo della Conoscenza per i dati di prova - Forced based.

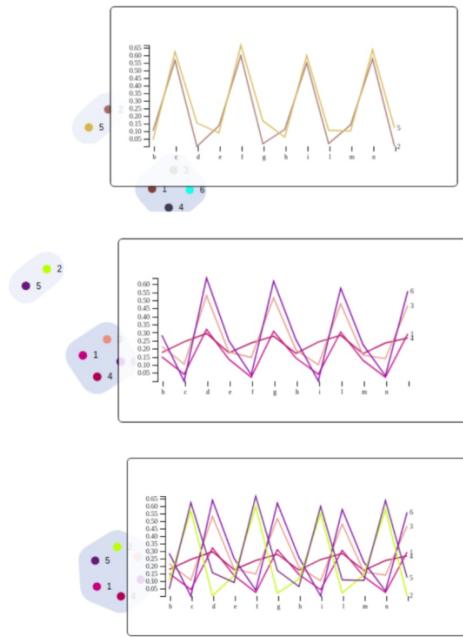


Figura 52: Reticolo della Conoscenza per i dati di prova - Statistiche.

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. Un esempio del fenomeno è illustrato di seguito.

La spiegazione del Reticolo malformato è da ricondurre alla natura stessa del 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 un discostamento causato dalla frequenza con il quale le domande vengono poste di cui l'addestramento della Rete ne risente troppo e che non riesce a mitigare del tutto; come si vede dalle figure seguenti.

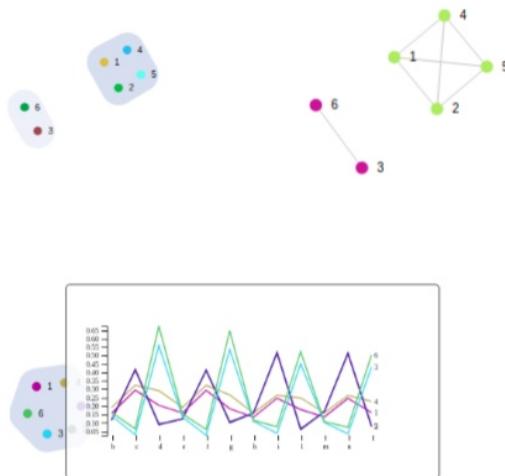


Figura 53: Reticolo della Conoscenza per i dati di prova - Presenza di valori di alterazione.

Le domande 6 e 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 Simple usato dall'applicativo per associare le domande, nel momento in cui correla tra loro due punti, il prossimo punto viene calcolato dal complessivo, invece che dal punto con una distanza appena superiore al punto appena Single. Quest'ultimo punto avrà un impatto superiore tanto più sono grandi le dimensioni del Reticolo. Per valutare al meglio l'impatto che di quanto affermato sopra, in fase di test mi sono preoccupata di analizzare e 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:

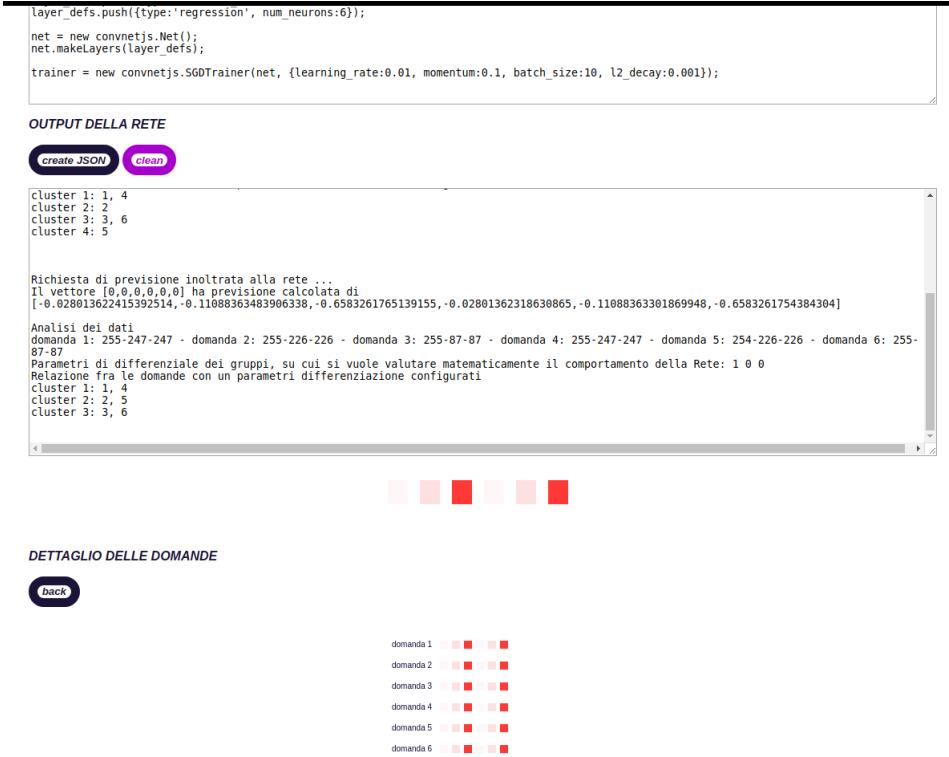


Figura 54: Risultati della Rete di prova con un set di dati puro (esclusivamente 1 e -1).

Come si può vedere dalla figura ?? 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. dalla Rete. Il fenomeno sorprendente, che mette in luce l'importanza di avere un valore 0 di risposta (non data) nel trainset, è mostrato nelle figure 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](#) [clean](#)

```

cluster 1: 1, 4
cluster 2: 2
cluster 3: 3, 6
cluster 4: 5

Richiesta di previsione inoltrata alla rete ...
Il vettore [0,0,0,0,0] ha previsione calcolata di
[-0.028013622415392514, -0.11088363483906338, -0.6583261765139155, -0.02801362318630865, -0.11088363301869948, -0.6583261754384304]

Analisi dei dati
domanda 1: 255-247-247 - domanda 2: 255-226-226 - domanda 3: 255-87-87 - domanda 4: 255-247-247 - domanda 5: 254-226-226 - domanda 6: 255-87-87
Parametri di differenziale dei gruppi, su cui si vuole valutare matematicamente il comportamento della Rete: 1 0 0
Relazione fra le domande con un parametri differenziazione configurati
cluster 1: 1, 4
cluster 2: 2, 5
cluster 3: 3, 6

```

DETtaglio delle domande

[back](#)

domanda 1	■	■	■
domanda 2	■	■	■
domanda 3	■	■	■
domanda 4	■	■	■
domanda 5	■	■	■
domanda 6	■	■	■

Figura 55: Risultati della Rete di test con un set di dati puro (esclusivamente 1 e -1), vettore previsione [-1, 0, 0, 0, 0, 0].

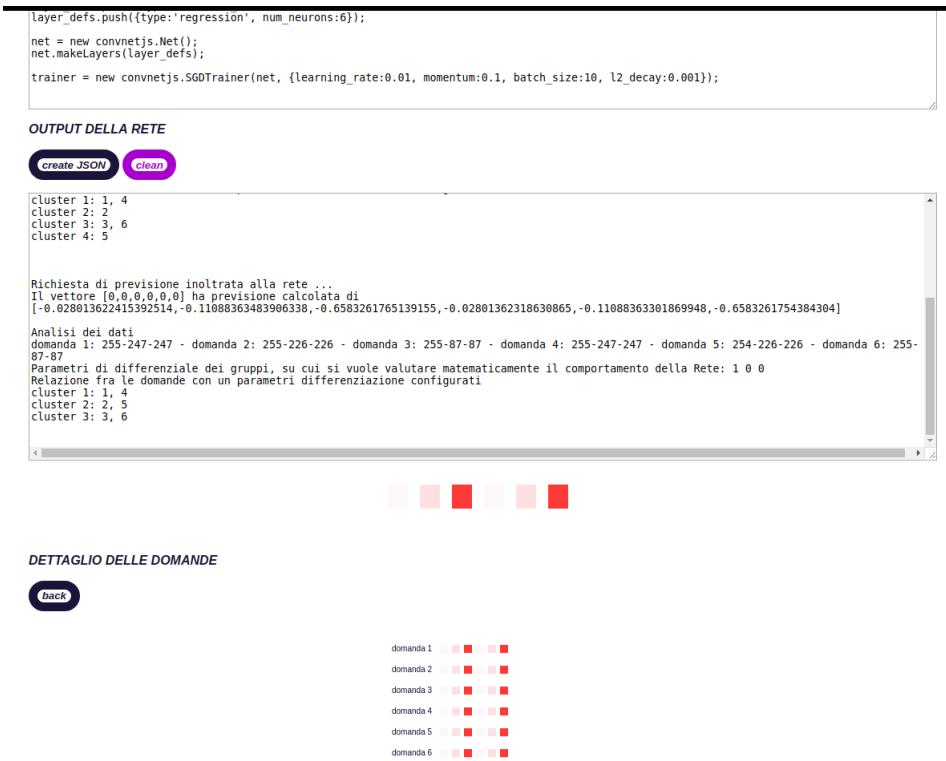


Figura 56: Risultati della Rete di test con un set di dati puro (esclusivamente 1 e -1), vettore previsione [0, 0, 0, -1, 0, 0].

Non avere un valore a 0 porta le le previsioni imposte alle singole domande, valutate in coppia, a mostrare come unica relazione stretta esistente la coppia coinvolta, scollegando ogni relazione con le altre coppie. 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 sulla previsione visiva della Rete, dal quale risulta impossibile la previsione di domande figlie e genitori.

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

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 dei set di dati anche in base alla loro frequenza; in modo da poterne valutare le differenze con la

Rete neurale e quali sono, se vi sono, 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.

Figura 57: Frequenza delle domande dei dati di test.

L'immagine sopra mostra 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 guida⁸. 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 di 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 si presenta 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;
 - la domanda 3 si correla con la domanda 6,
 - la domanda 2 con la domanda 5.

⁸con impiego del Grafo della Conoscenza (mostrato in figura 6)

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 esistente tra le domande accoppiate⁹.

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 dell'occorrenza (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'impiego 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

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

⁹per i dati di prova le domande accoppiate sono (1;4), (3;6) e (2;5)

(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.
-

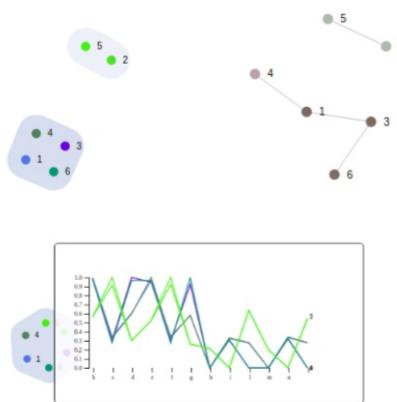


Figura 58: Reticolo della Conoscenza per i dati di prova - Frequenza.

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, ricalcano fedelmente quanto dichiarato dalla matrice correlazione ottenuta dal modello generato dall'uso della PCA, e visibile dall'immagine 40. 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 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 se si tratti di una valutazione per frequenza, che di previsioni ottenute dalla Rete neurale, il Reticolo tende ad ogni nuovo set di dati, 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 alcuna 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.

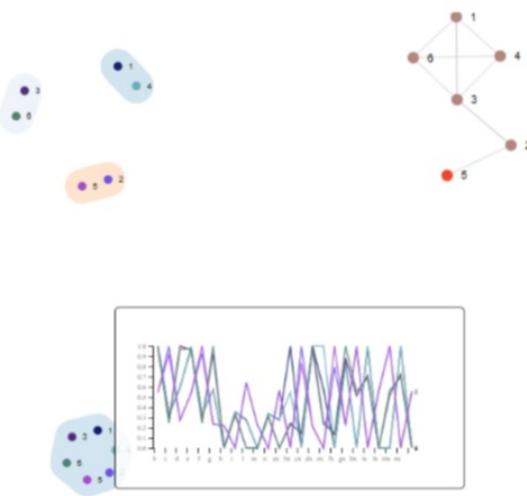


Figura 59: Reticolo della Conoscenza per i dati di prova con indovinato - Frequenza.

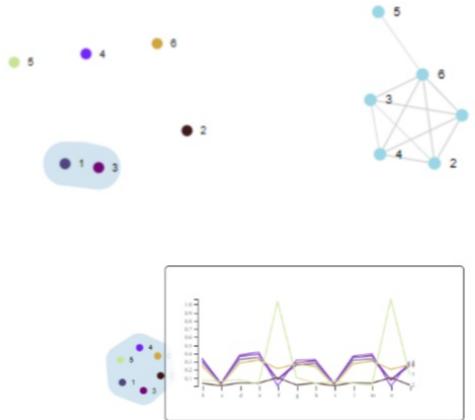


Figura 60: Reticolo della Conoscenza per i dati di prova con indovinato - Previsioni.

4.4 Creazione del Reticolo della Conoscenza sui dati delle domande del database

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

1. Studio del testo delle domande per individuare quali sono i gruppi di argomenti;
2. Formulazione 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
aelkshwnzj	Nessuno
bajsownqty	Nessuno
bdwjgxtdsm	Nessuno
bfmfxcudzq	Nessuno
bkfstjjeaw	Nessuno
bojihohtuw	Nessuno
ccsofpywab	Eulero-Venn
cfkjhtcjjp	Nessuno
cfxigdfgsu	Eulero-Venn
cidpdlimnu	Nessuno
cxjganqls	Nessuno
cxwrijrltfv	Nessuno
diucioyoxo	Nessuno
dpzaqasqro	Nessuno
ectvbesohs	Nessuno
efbcntiwcr	Nessuno
elbottcdse	Nessuno
fbekkymdja	Nessuno
fclxhnnodej	Nessuno
flsdkdhwpz	Nessuno
fpiajgnnfb	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
jwvreihjvv	Nessuno
jygnlagbhv	Nessuno
kkhcmynavp	Nessuno
kprsxjkubz	Serie numeriche
czorhoohrb	Nessuno
lhuxxlqtyz	Nessuno

ljypeljfkhh	Eulero-Venn
lqyvhqfmqd	Nessuno
lrqrblchp	Nessuno
mfycpmjyzc	Nessuno
mhairflmrt	Eulero-Venn
mnrnuysydx	Serie numeriche
ngtquzulpk	Nessuno
nrawevuhkv	Nessuno
oaszeadbqh	Nessuno
obwqtgphek	Eulero-Venn
okklvldqlc	Nessuno
osvluiqhbp	Nessuno
pfctjunpxs	Serie numeriche
pfsqvqgsfen	Serie numeriche
poiuytrewq	Eulero-Venn
ptfkuubppy	Nessuno
ptwccskjcf	Serie numeriche
pxvpfrjqnb	Serie numeriche
qbnjsxmnxz	Serie numeriche
qfxndoiusk	Nessuno
qmzlhizqng	Nessuno
qsqbqjczyv	Eulero-Venn
ribhqompdp	Nessuno
sucpsfikqd	Nessuno
tnmucbefiy	Nessuno
toctsuegmj	Nessuno
tpfgrbtdu	Nessuno
ukafbvkjxl	Nessuno
ungslccrdw	Serie numeriche
uoohruqjig	Serie numeriche
uoztprdkre	Nessuno
upmqfdzqqd	Nessuno
voqwkjnxuv	Eulero-Venn
vtjiowgfp	Nessuno
wfthuqobvv	Nessuno
wpfhnhkztw	Nessuno
wqkhvmdpol	Serie numeriche
xmqopkamew	Serie numeriche
xnyfhvvlzx	Serie numeriche
xvbulkwvol	Nessuno
yxkoznnxis	Nessuno
zcflqcwpgh	Nessuno
zegfnrndwr	Nessuno
zptjzdwilf	Nessuno

zqrfurehmz	Serie numeriche
zrehhcjyl	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 della sezione 2 la scelta dell'architettura adatta per una rete che deve trattare una grande mola e di cui non si ha piena conoscenza del risultato finale che si deve ottenere ¹⁰ rende tale compito molto arduo.

L'unico modo di affrontare il problema è procedere con l'analisi del set di dati in input aiutandosi 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.SGDTTrainer(net, {learning_rate:0.01,
momentum:0.1, batch_size:10, 12_decay:0.001});
```

Per formularla 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 del caso in analisi. Non mi aspetto abbia dei risultati accesi, 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});
```

¹⁰perché soggetti all'indovinato come visto in ??

```

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});

• 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});

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});

• 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:

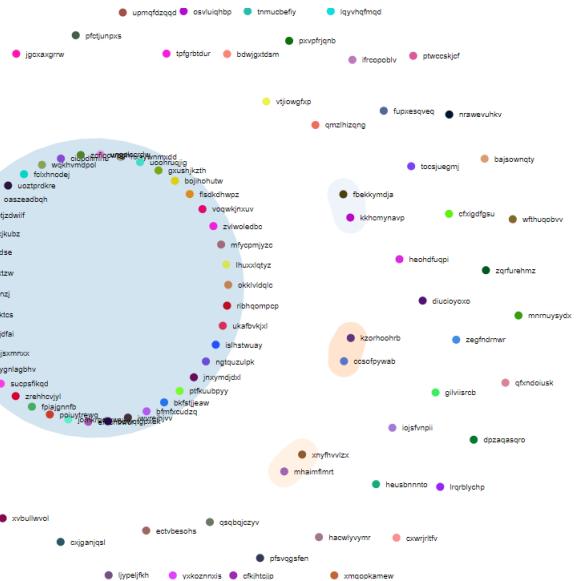


Figura 61: Reticolo della Conoscenza layers(12, 12) - Cluster Based.



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

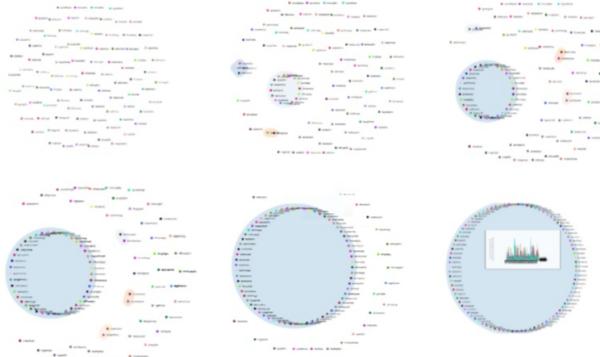


Figura 63: Sequenza di creazione del Reticolo della Conoscenza layers(12, 12) - Cluster Based.

Osservazioni I risultati ottenuti sono buoni ma non eccelsi. Prima di tutto si crea un unico cluster, e non due come attendo, però in questo si vanno a 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	obwqtgpxek
qbhjsxmnx	poiuytrewq
ungslccrdw	voqwkjnxuv
uoohruqjig	zvlwoledbc
wqkhvmdpol	
zrehhcvjvl	

Tabella 6: Domande contenute nel primo cluster

Le domande correlate nei cluster (*mhaimflmrt*, *xnyfhvvlxz*), (*ccsofpywab*, *kzorhoohrb*) e (*fbekkymdja*, *kkhcmynavp*), visibili in figura 61 non presentano alcuna somiglianza di tematica.

Riducendo la correlazione tra i cluster si generano 2 cluster, come visibili dall'immagine 63; ma non viene rispettata la differenziazione tra domande di 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 2.2.3: 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 64: Reticolo della Conoscenza layers(4, 4, 10) - Cluster Based.

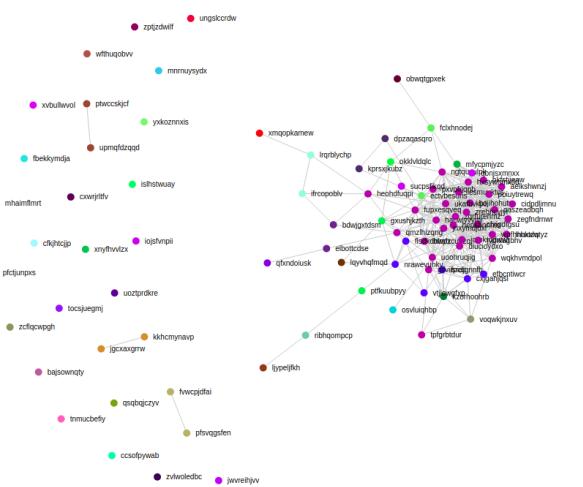


Figura 65: Reticolo della Conoscenza layers(4, 4, 10) - Forced Based.

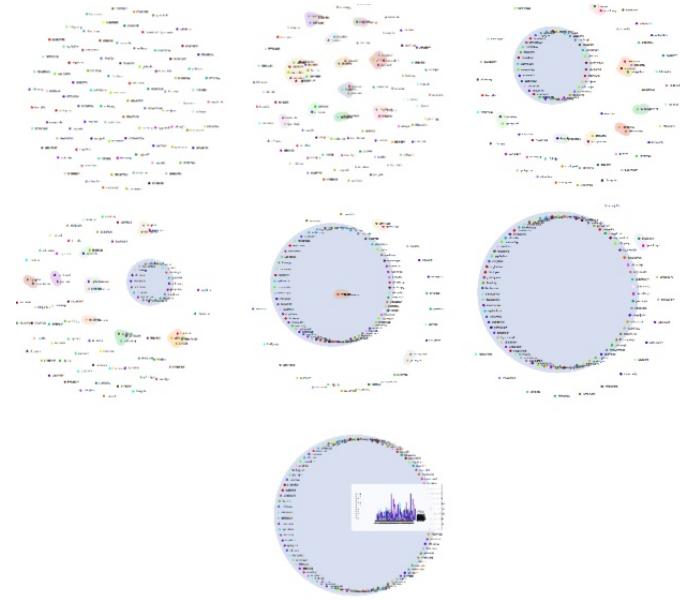


Figura 66: Sequenza di creazione del Reticolo della Conoscenza layers(4, 4, 10) - Cluster Based.

Osservazioni Questo, ha mio avviso è, uno dei Reticolo più interessanti che ho generato, riguardante il set di domande di logica. Esiste un cluster principale e una decina di gruppi di dimensione via via sempre minore. A mano a mano che viene inalzata la threshold, i gruppi vengono inglobati dal gruppo principale. I punti isolati sono in numero notevole, come si può vedere dall'immagine 64 anche se la threeshold di competenza è quasi a metà.

I punti isolati rappresentano, per la maggioranza dei casi tutte quelle domande che non trovano correlazione nè tra le serie numeriche, nè tra i diagrammi di Eulero-Venn. Nel cluster di dimensione maggiore trovano collocazione i seguenti punti:

Serie Numeriche	Eulero-Venn
zrehhcvjyl	cfxigdfgsu poiuytrewq

Tabella 7: Domande contenute nel primo cluster (blu)

Ci sono tuttavia anche altri cluster interessanti:

Serie Numeriche	Eulero-Venn
wqkhmdpol	gilviisrcb

Tabella 8: Domande contenute nel secondo cluster (rosetta)

Serie Numeriche	Eulero-Venn
zqrtfurehmz uoohruqjiq zqrtfurehmz	

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

Si può definire come il primo e il secondo cluster hanno una predominanza delle domande, per quanto concerne le domande catalogate, per le domande 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 configurazione a layers (4, 4, 10). Di seguito riporto esclusivamente i risultati ottenuti dalla Generazione del Reticolo.



Figura 67: Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.



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

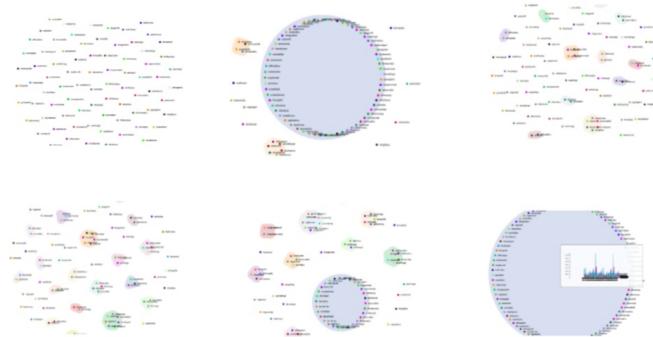


Figura 69: Sequenza di creazione del Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.

Quarta architettura: layers(4, 4, 5, 5) Il Reticolo della Conoscenza generato è il seguente:

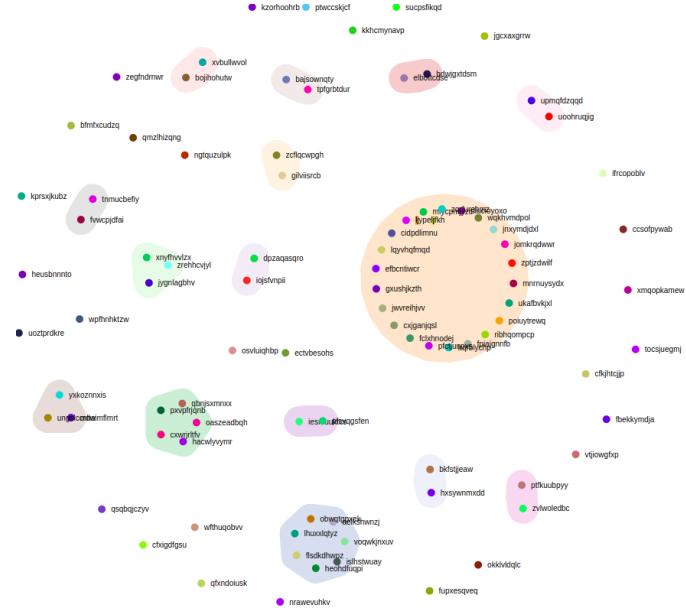


Figura 70: Reticolo della Conoscenza layers(4, 4, 5, 5) - Cluster Based.



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

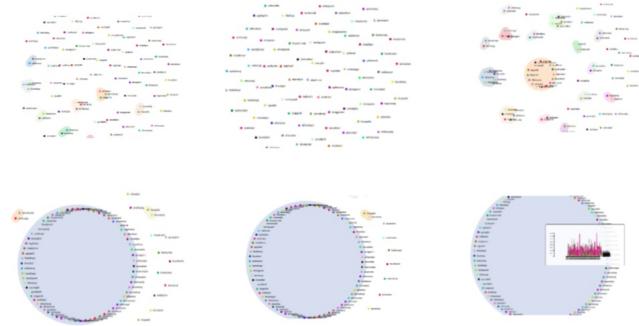


Figura 72: Sequenza di creazione del Reticolo della Conoscenza layers(4, 4, 5, 5) - Cluster Based.

Osservazioni 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 i 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 presentano strettamente correlate. Tale architettura porta dei benefici inferiori rispetto a quella descritta in precedenza; ma ho comunque ritenuto fosse interessante da mostrare. Di seguito descrivo solo alcuni cluster che ritengo significativi.

Serie numeriche	Eulero Venn
pfctjunpxs	ljypeljfhk
zqrfurehmz	
wqkhvmdpol	poiuytrewq

Tabella 10: Domande contenute nel primo cluster

poiuytrewq e *ljypeljfhk* derivano dal medesimo cluster, separati dalle domande sulle serie. Come si può vedere dalle immagini in figura 72 alcuni cluster di dimensione ridotta, contengono mini gruppi di domande sulle serie:

- *zrehhcwjyl* 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 domae di logica

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



Figura 73: Sequenza di creazione del Reticolo della Conoscenza layers(4, 4, 10) con Frequenza - Cluster Based.

In questo caso 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 che un utente ha di indovinare e dalla presenza di risposte non date.

Come si può vedere dall'immagine sopra sono individuabili due cluster principali, composti:

cluster 2 (blu scuro)	cluster 3 (blu chiaro)
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 que-

sto 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, della Rete, 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 Domande SQL

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, contenute all'interno del database aziendale, in modo da ottenere un poter studiare come il Reticolo si comporta con una variazione del trainset.

Le architetture scelte hanno tenuto conto quanto risultato durante la costruzione del Reticolo 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 per estrapolare i dati, erano presenti anche delle domande di Informatica Generale. E' dunque lecito ipotizzare che le domande di Informatica Generale compongano 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
akczmgrpdbq	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
dkzgokqmiq	SQL
dlfyuvcbqx	SQL
dllpxhzfkt	SQL
dzxsvydlth	SQL
ebrpcrnecd	SQL
eftrkdthng	SQL
elqzybkpj	SQL
enitorndkz	Informatica Generale
enmkdvzmyk	SQL
ewplljreqn	Informatica Generale
ffbcwwtnzk	SQL
fhqcltuqw1	SQL
fjmpjvualm	SQL
fkdgytnwzf	SQL
fumfugvkrl	Informatica Generale
fuzfbwlutx	SQL
fyrpicjcan	SQL
fzoxjumtyn	SQL
gtxwazgyqz	SQL
guatssqajf	SQL
gyerxpmgqa	SQL
hntvpqnrc	Informatica Generale
hyfoktjlpp	SQL
idfxvfzprt	SQL
ieimtqtzpk	SQL
iezewxpthv	SQL
ifrvxoopny	SQL
ifvdeoyre	SQL
ikyqsdhuit	SQL
jbpjxhaxyy	SQL
jedvsptntk	SQL
jycipstmyp	SQL
kbnwlxqtqc	SQL
kexorxwipq	SQL
kfbpwmrkgz	SQL
kjecbikedc	SQL
kqkjrygfsc	SQL
kzzvvwynxn	SQL
lgiqayejf	SQL
ljozpygnua	SQL
lohsalpqn	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
pnornaipiy	Informatica Generale
pozlktnbwi	Informatica Generale
pralrcttdk	Informatica Generale
ptfyswhhoi	SQL
ptvnijvlyr	SQL
qfpxivqgfs	SQL
qhpmwyunln	SQL
qhurfzksbi	SQL
qyfsksfpqr	SQL
rcvjbkxjb	Informatica Generale
rfveenepae	SQL
rjqonwgagd	Informatica Generale
rjrbsovupf	SQL
rkpyhnweoz	Informatica Generale
rlbnjnuraa	SQL
rqlhmwvvfr	SQL
shwidpusut	SQL
slfnwkouex	SQL
soaxrpkmix	SQL
tahwadyybr	SQL
tmalpukvph	SQL
tmptxxfqlg	SQL
twjbzfuzhs	SQL
uciubnhesv	SQL
uctsvufnre	Informatica Generale
ukteiawtcu	SQL
uoujmoastk	SQL
usloxptlnv	SQL
vcetherudx	SQL
vvesvpdexu	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
zmivrwbccja	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:

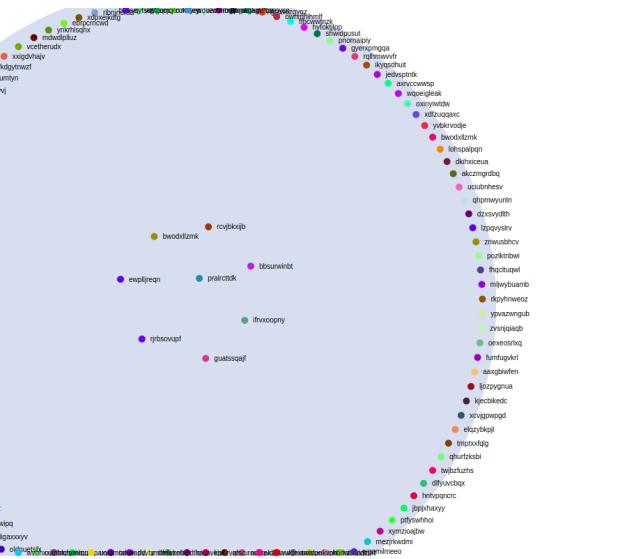


Figura 74: Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.



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

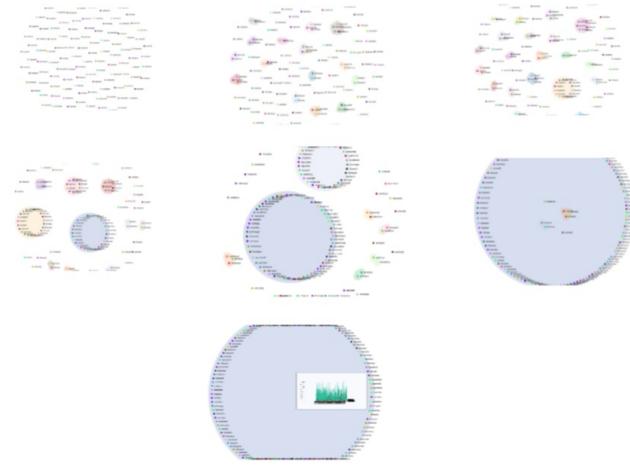


Figura 76: Sequenza di creazione del Reticolo della Conoscenza layers(2, 4, 10) - Cluster Based.

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

Codice	Argomento
ifrvxoopny	SQL
pralrcttdk	Informatica Generale
rcvjbkxijb	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:

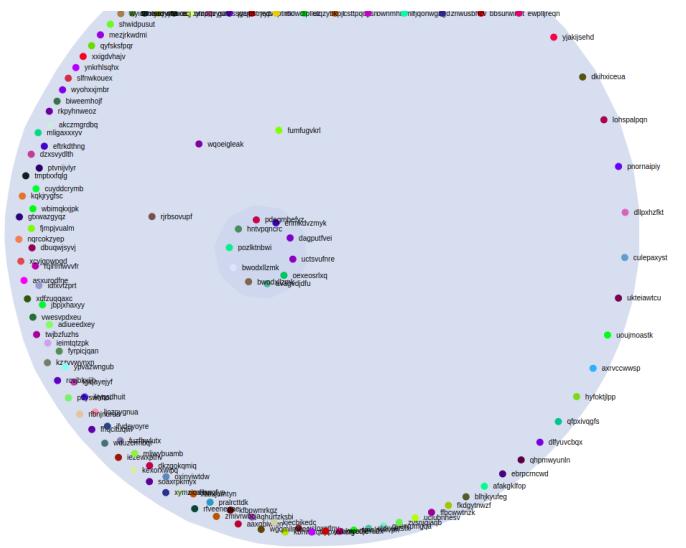


Figura 77: Reticolo della Conoscenza layers(2, 2, 10) - Cluster Based.

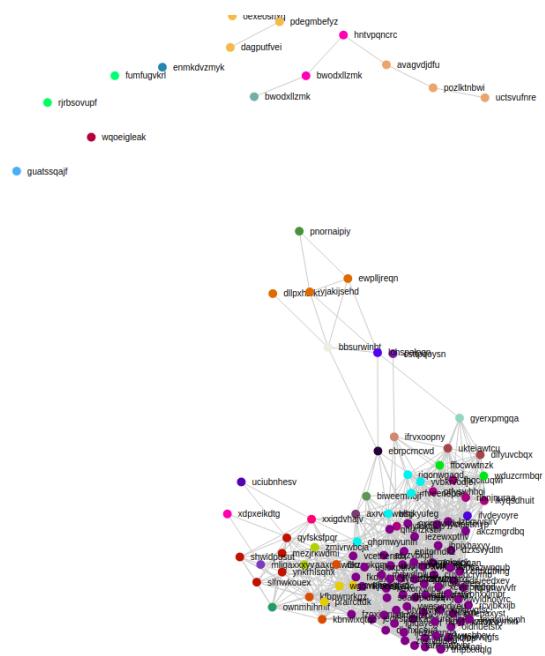


Figura 78: Reticolo della Conoscenza layers(2, 2, 10) - Forced Based.

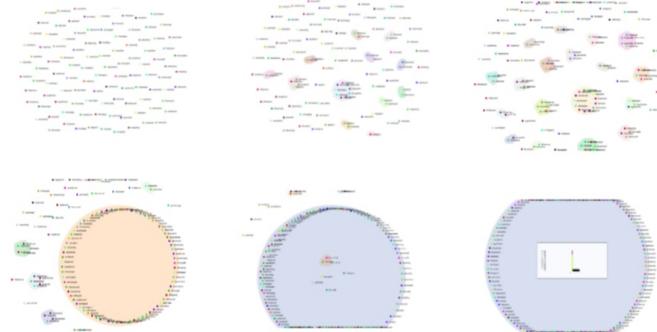


Figura 79: Sequenza di creazione del Reticolo della Conoscenza layers(2, 2, 10) - Cluster Based.

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

Codice	Argomento
enmkdvzmy	Informatica Generale
uctssvufnre	Informatica Generale
pdegbefy	Informatica Generale
pozikthbw1	Informatica Generale
oexeosrxq	Informatica Generale
hntvpqnrc	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, dal quale ne risultano estranee solo le domande *bowdxllzmk* e *avagvdjdfu*.

Inoltre vengono associati in relazione stretta:

Codice	Argomento	Trattazione del quesito
oldnuetsfx fyrpicjqan	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 di logica.

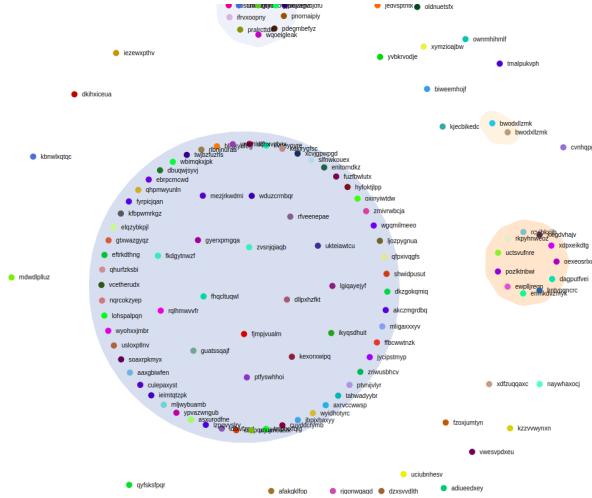


Figura 80: Sequenza di creazione del Reticolo della Conoscenza layers(2, 2, 10) con Frequenza - Cluster Based.

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 logica 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 in gruppi assieme a domande sql e rientrano fin da subito all'interno del cluster principale.

Come pro di questa tecnica è che individua subito che i dati contengono, erroneamente, un a domanda inserita due volte *bwodxlzmk*.

cluster 2 (rosa)	cluster 3 (azzurro)
pozltnbw	wqoeigleak
ewlppljreqn	pralrcttdk
hntvpqncrc	pnomaipjy
oexeosrxq	fumfugvkr
rcvjldkxib	rjrbsovupf
rkpyhnweo	

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

Inoltre le domande che trattano la sargabilità (*oldnuetsfx*, *fyrpiecgan*, *fuzfbwlutx*), o meno di un predicato non vengono messo in nessun modo in correlazione, che invece la Rete neurale è in grado di individuare.

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.

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 browser¹¹ per questo ho deciso di omettere le keywords exports come preambolo dei metodi soggetti a test; tale aggiunta è riservata all'utente nel momento precedente al test running.

¹¹Il 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 81: Resoconto unit test con successo.

5.2 Sistema operativo e browser

Il progetto è stato sviluppato su sistema operativo Linux e ripetutamente testato per tutta la durata dello stage sul sistema operativo Windows messo a disposizione in ufficio. L'applicativo della Rete neurale è sviluppato ad doc per browser Chrome in cui il caricamento di grandi moli di dati ha prestazioni migliori. Sconsiglio l'utilizzo della Rete neurale in uso per i dati del database su Microsoft Edge e Firefox , in quanto il caricamento del file di allenamento è molto lento.

5.2.1 Specifiche di sviluppo

- Sistemi operativi di sviluppo:

- 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