

Indice

1	Analisi dei dati di probabilità	3
1.1	Problema in esame	3
1.2	Caratteristiche degli eventi di coppia	3
1.2.1	Eventi indipendenti	4
1.2.2	Eventi dipendenti	4
1.3	Evento conosciuto ed evento indovinato	5
1.3.1	Probabilità di rispondere correttamente ad una domanda	6
1.3.2	Il piano	8
2	Rete neurale	13
2.1	Test effettuati	14
2.1.1	Architettura della rete: 4 neuroni per ciascuno dei 2 layers	15
2.1.2	Architettura della rete a 2 neuroni per ciascuno dei 2 layers	16
2.1.3	Architettura della rete a 4 neuroni per 1 layer	19
2.2	Sviluppo della rete delle domande nel database aziendale	20
2.2.1	Montaggio e configurazione della rete	20
2.2.2	Natura delle domande contenute nel database aziendale	22
2.2.3	Test e Documentazione	23
2.3	Funzionalità offerte dall'interfaccia della Rete neurale delle domande nel database	40
2.3.1	Diagrammi dei casi d'uso	41
3	Principal Component Analysis	61
3.1	Metodologia applicata	61
3.2	Sviluppo	62
3.3	Risultati ottenuti	64
3.3.1	Calcolo della Principal Component Analysis	64
3.3.2	Calcolo degli autovettori ed individuazione di Summary	65
3.3.3	Calcolo degli autovettori	73
3.3.4	Calcolo della matrice di correlazione	73
3.4	Conclusione dell'analisi	75
4	Costruzione del Reticolo della Conoscenza	78
4.1	Descrizione del sistema	78
4.1.1	Configurazione	79
5	Creazione dei file CSV	80

6	Creazione del Reticolo della Conoscenza per sui dati di Prova	80
6.0.1	Osservazioni	81
6.0.2	Osservazioni Reticolo dati di prova	81
7	Creazione del Reticolo della Conoscenza sui dati delle domande nel database	81
7.1	Osservazioni	82
7.1.1	Osservazioni Reticolo dati del database	82

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:

AB

\wedge

$A \ B$

\vee

Z

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.

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

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.

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

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 per cui 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.

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

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 stanziava 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:

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

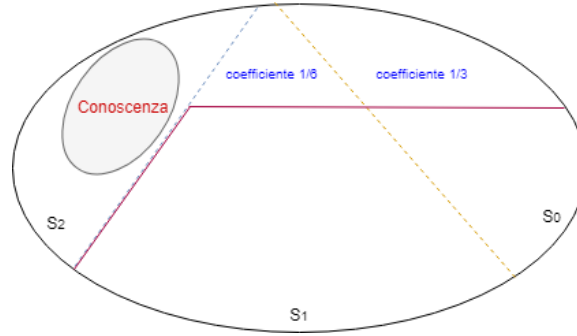


Figura 1: Rappresentazione insiemistica della probabilità di rispondere correttamente ad una domanda: $P(A)$

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 appropria 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).

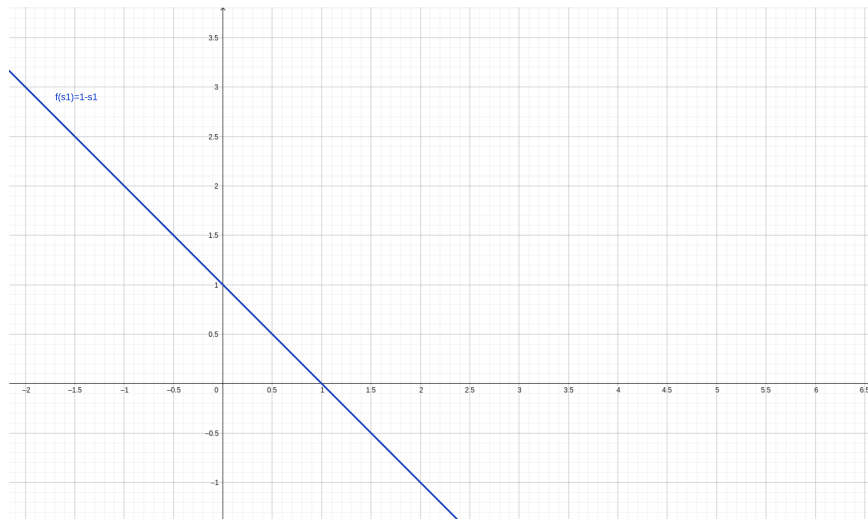


Figura 2: Rappresentazione della retta passante per $S_0 = 0$

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 variabili coinvolte) e mediante la retta $S_1 = -4S_2 - 2 + 6F$ (che permette di calcolare il fascio di rette tangenti alla prima retta).

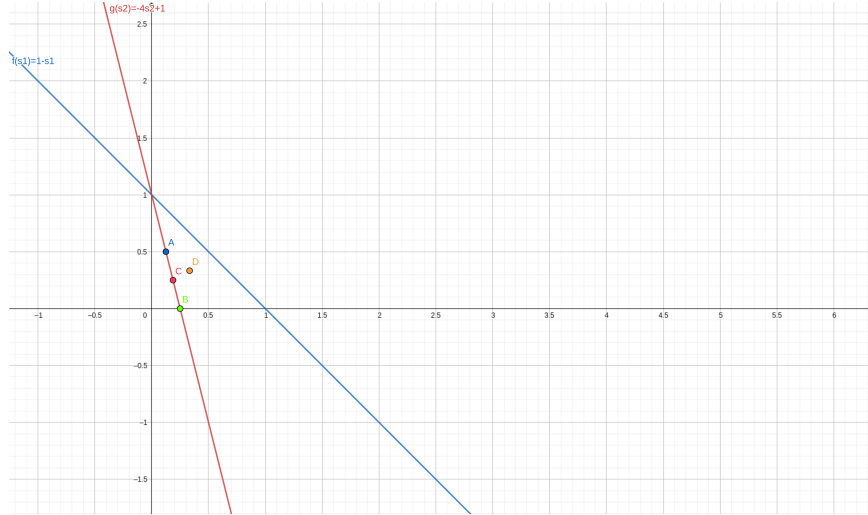


Figura 3: Rappresentazione di P(A) per una frequenza 0.5 proiettata su assi $S_0 = 0$, S_1 e S_2 .

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):

$$S_1 = 0.5 = \frac{1}{2}$$

$$\frac{1}{2} = -4S_2 + 1 \rightarrow 4S_2 = 1 - \frac{1}{2} \rightarrow S_2 = \frac{1}{8}$$

$$S_0 = 1 - \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$$

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):

$$S_1 = 0$$

$$S_2 = \frac{1}{4}$$

$$S_0 = 1 - \frac{1}{4} = \frac{3}{4}$$

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):

$$S_1 = \frac{1}{4}$$

$$S_2 = \frac{3}{16}$$

$$S_0 = 1 - \frac{1}{4} - \frac{3}{16} = \frac{9}{16}$$

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):

$$S_1 = \frac{1}{3}$$

$$S_2 = \frac{1}{3}$$

$$S_0 = 1 - \frac{1}{3} - \frac{1}{3} = \frac{1}{3}$$

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 nè una nè due risposte, per cui 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 stata la semplicità nell'utilizzo del linguaggio javascript; l'aspetto negativo ha riguardato la totale mancanza di mantenibilità della libreria stessa che comporta la scarsità di esempi applicativi, oltre alla documentazione ufficiale, che costringono lo sviluppatore ad una ricerca approfondita personale in un ambiente ove le nozioni si presentano scarse e a continue prove per verificare la validità del codice prodotto.

Questions test - Prevision Neural Net

CONFIGURAZIONE RETE NEURALE

```
layer_defs = [];  
layer_defs.push({type:'input', out_sz: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 inoltrata alla rete ...  
Richiesta di previsione accettata  
Il vettore [0,1,0,0,0,0] ha previsione calcolata di  
[-0.14178025275470646,-0.014639237878831693,-0.00131295482446743,0.11790470015357472,0.14410273493990078,0.1329311955165462]  
Rete Neurale in attesa ... Inserire risultati del test...  
Inserimento risposte alle domande del test andato a buon fine  
Ricapitolazione dati inseriti: Risposte ottenute [0,-1,0,0,0,0]  
Inizio allenamento della rete ...  
Richiesta di previsione inoltrata alla rete ...  
Richiesta di previsione accettata  
Il vettore [0,1,1,0,0,0] ha previsione calcolata di  
[-0.9703841414997074,-0.7824502014273779,0.34748891066895563,-0.8448991676224333,-0.5825080715009475,0.29369850449980084]  
Rete Neurale in attesa ...
```

INSERIMENTO DATI DI ADDESTRAMENTO

Number of fields (max 9999): [Insert fields](#)

Input1 Input2 Input3 Input4 Input5 Input6

[save](#)

[start trainer](#)

Vettore di previsione

[prevision trainer](#)

Figura 5: 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 mi sono occupata di implementare 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 che ho 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

Alcune decisioni che ho preso durante la scelta dell'architettura della rete riguardano i seguenti settori:

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, in output, anche in virgola mobile e composto di tanti elementi quanti sono i neuroni di regressione dichiarati. Per la mia rete di prova è necessario dichiarare 6 neuroni in regressione perchè l'output, appunto, che ci si aspetta dal sistema è di 6 elementi.
3. Per costruire un dataset di dati consistente che permettesse alla rete di imparare qualcosa ho costruito un grafo della conoscenza con lo scopo di mettere in relazione degli argomenti che coinvolgono uno o più domande.

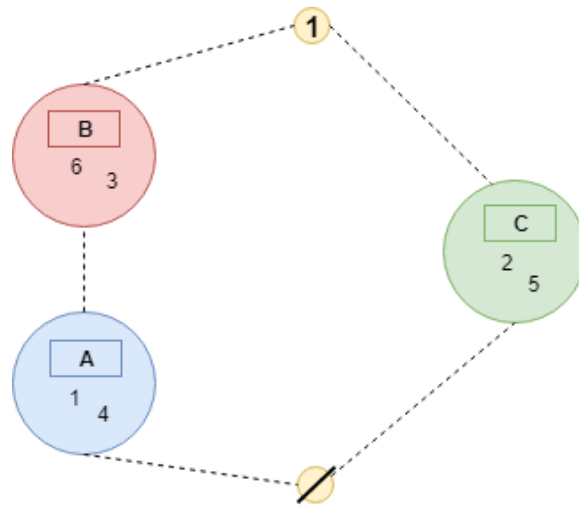


Figura 6: Grafo rappresentante le relazioni esistenti tra il set di domande di prova.

Per svolgere l'apprendimento ogni vettore, facente parte del dataset, viene dato in pasto alla rete che a sua volta provvede alla sua assimilazione come conoscenza mediante la 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.

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; mol-
to 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 coppie 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; conseguenza diretta di un numero eccessivo di neuroni dati alla rete per l'apprendimento rispetto al training set svolto, generatrice di una situazione 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 overfitting, 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 aziendale

2.2.1 Montaggio e configurazione della rete

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. 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 presenta anche in versione ridotta.

Per rendere più comprensibile le previsioni di probabilità ottenute, a seguito dell'addestramento della rete e della data in pasto del vettore previsione, ho realizzato un'immagine canvas in cui ogni domanda viene raffigurata con un quadrattino colorato in base alla previsione risultante (verde se a 1, bianco a 0, rosso a -1, gradazioni di bianco - verde e bianco - rosso per i valori intermedi).

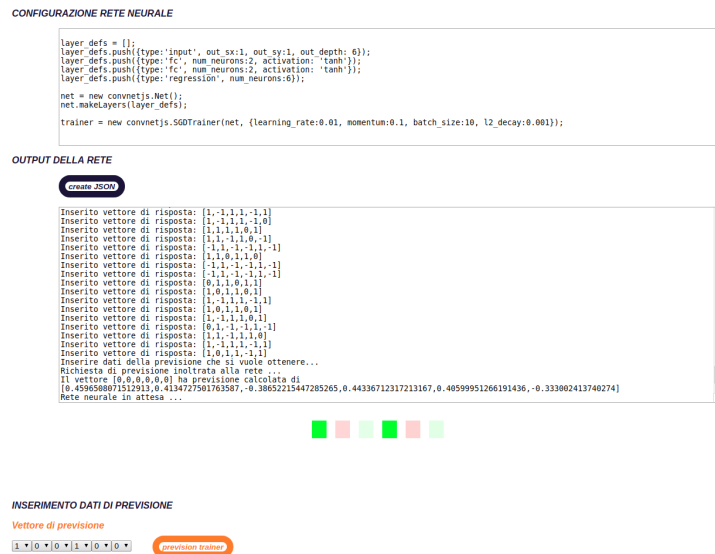


Figura 7: Rete di prova dopo lo sviluppo del canvas per le previsioni.

- Solo una piccola parte delle domande 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. Quello che mi sembra opportuno ricercare testando l'architettura di della rete è che si vadano a formare dei cluster.

2.2.3 Test e Documentazione

Durante la settimana dal 10/06 al 18/06 ho effettuato quanto definito al 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]

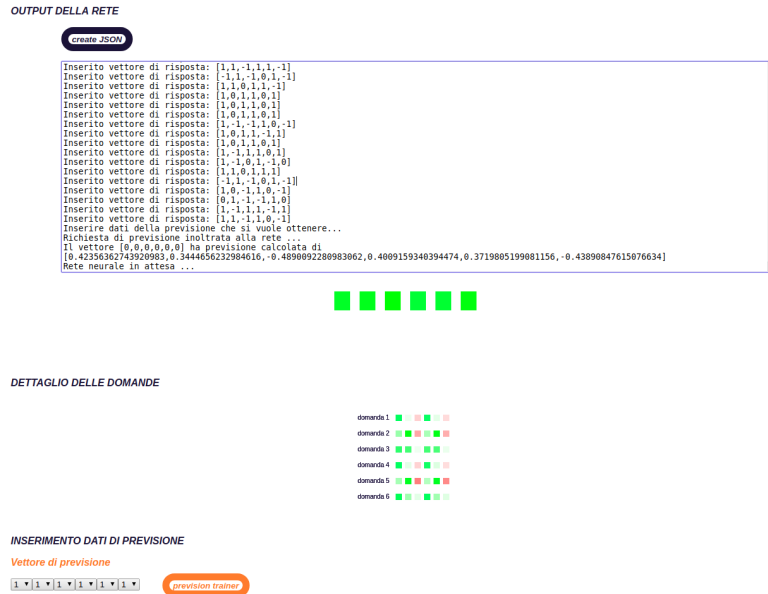


Figura 9: Risultato della rete di prova a seguito di un vettore di previsione $[1, 1, 1, 1, 1, 1]$ in input.

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 simile 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 con 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 simile 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) ma 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% o con una differenza massima di 5 punti colore rispetto al canvas risultante per ogni domanda.

Tuttavia non sono, durante questo periodo, riuscita ad individuare un'architettura sufficientemente stabile per prevedere risultati attendibili; a causa della molteplicità di dati che hanno aumentato esponenzialmente la complessità di analisi rispetto alla Rete di prova. Tale complessità è rimasta costante anche successivamente allo studio del contenuto delle domande nel database aziendale. Architettura della rete utilizzata:

- `layer_defs = []`;

```

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

```

```

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

```

```

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

```

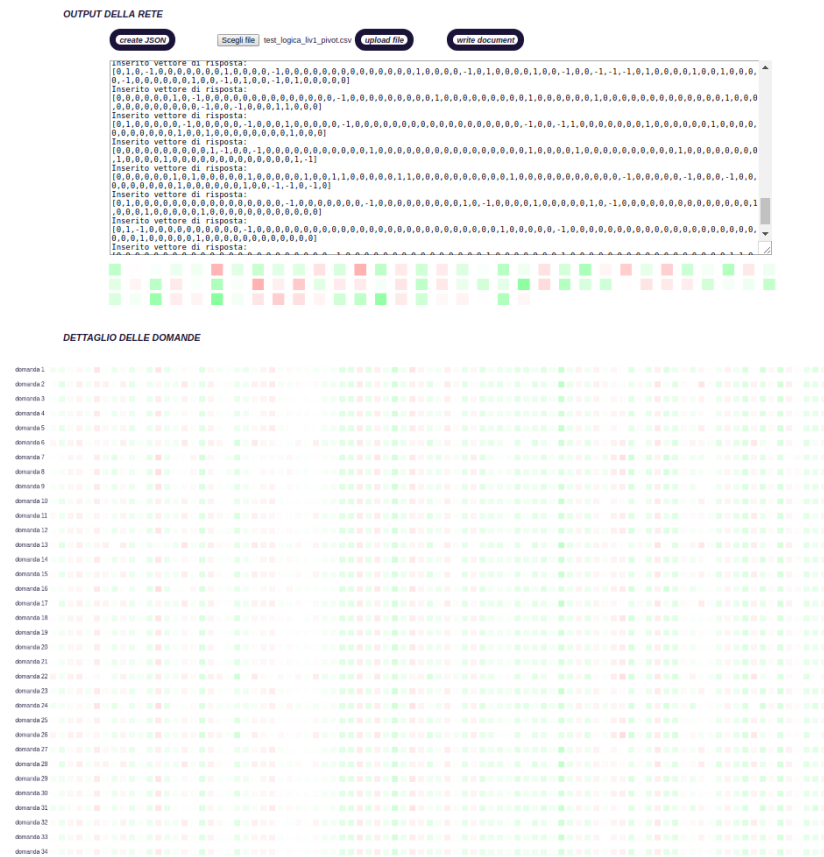


Figura 11: Risultato della rete del database a seguito di un vettore di previsione [1, 1, 1, 1, 1, 1] in input.

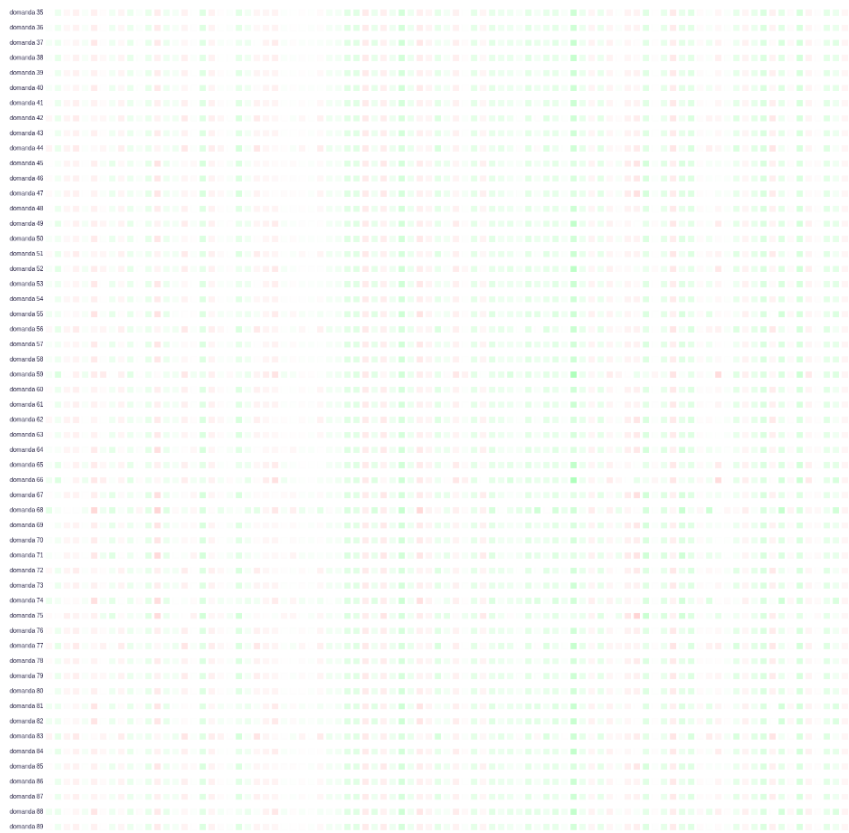


Figura 12: Risultato della rete del database a seguito di un vettore di previsione $[1, 1, 1, 1, 1, 1]$ in input.

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

1. Analisi verticale:

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

2. Analisi orizzontale: Appaiono in relazione stretta 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.

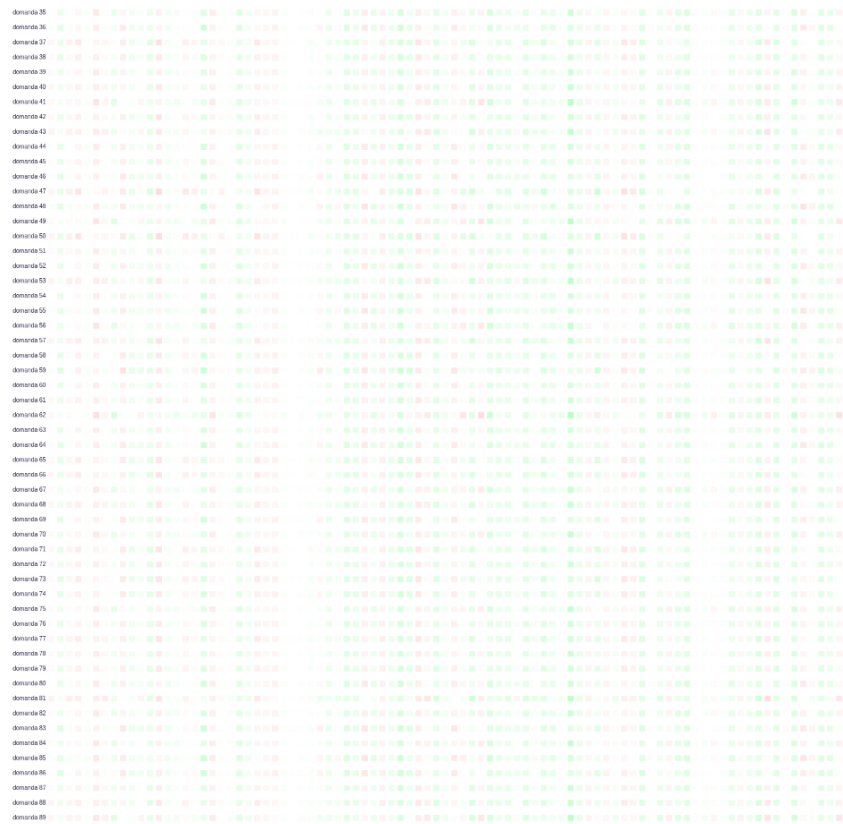


Figura 14: Risultato della rete del database a seguito di un vettore di previsione $[-1, -1, -1, -1, -1, -1]$ in input.

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 in relazione stretta 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.

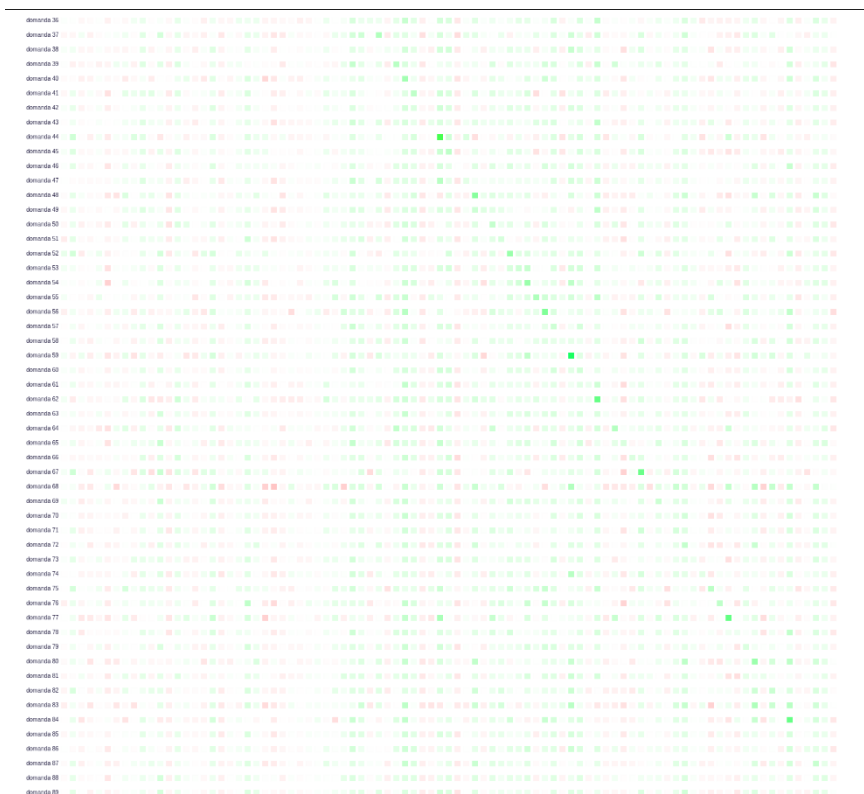


Figura 16: Risultato della rete del database a seguito di un vettore di previsione [1, 1, 1, 1, 1, 1] in input.

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.

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 relazione stretta le seguenti domande:

– Vengono meno le relazioni individuate precedentemente.

```
• 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});  
  
• 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});
```

L'obiettivo dei miei test è stato individuare una configurazione abbastanza "forte" da permettere il riscontro di correlazioni e dei possibili cluster esistenti con il solo ausilio della Rete. Tuttavia per proseguire il mio studio è indispensabile che mi avvalga di anche altri strumenti di supporto. Il prossimo passo è costruire un modello di dati mediante la tecnica di **Principal Component Analysis**.

2.3 Funzionalità offerte dall'interfaccia della Rete neurale delle domande nel database

La Rete neurale sviluppata è stata pensata per una persona che deve svolgere analisi dei dati mirata alla previsione dei risultati su un caso studio specifico (la dimensione è fissata a 89). 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 dell'architettura in uso, con indicazione della tipologia di trainer utilizzato. Questa sezione permette la modifica delle variabili di configurazione¹;

¹Ogni modifica del numero di neuroni espressi nella regressione deve provocare la modifica del file di previsione 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 si occupano di interagire 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

UC0_g: Operazioni utente - Visione generale interfacciamento con la Rete neurale

Per ogni funzionalità espressa in UC0_g è stato creato un diagramma differente in modo da evitare connessioni tra le frecce e una corretta visualizzazione delle estensioni.

- **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 ha potuto effettuare analisi e osservazioni sui dati di interesse;
- **Scenario principale:**
 1. L'utente può effettuare il caricamento dei dati di analisi nella Rete in formato txt o csv (UC1, UC2, UC3);
 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 parametri di previsione (UC8);
 7. L'utente può visualizzare i risultati della Rete ottenuti dai dati caricati e a seguito delle operazioni di previsione (UC9);
 8. L'utente può visualizzare il contenuto della Rete addestrata in formato JSON (UC10);
 9. L'utente può procedere all'eliminazione dello storico della Rete contenuto all'interno della textarea dedicata (UC11).
- **Estensioni:**
 1. Il browser in uso dall'utente non supporta lo standard HTML. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC9);
 2. File caricato dall'utente nella rete non conforme al formato di dati txt o csv richiesto come prerequisito dalla Rete. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC10).

UC1: Caricamento formato dati

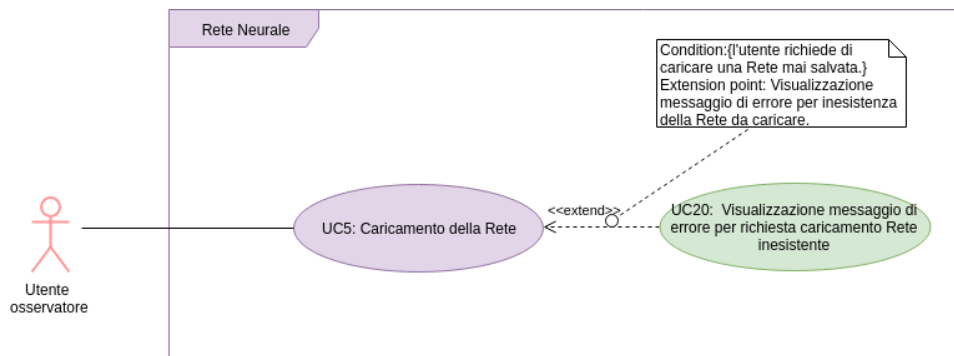


Figura 17: UC0_g: Operazioni utente - Visione generale interfacciamento con la Rete neurale (Caricamento formato dati).

- **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 nel formato idoneo;
- **Scenario principale:**
 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).
- **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 (UC9);
 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 (UC10).

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:**
 1. L'utente sceglie il file di dati che vuole caricare nella Rete;
 2. L'utente confermare l'operazione.

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:**
 1. L'utente sceglie il file di dati che vuole caricare nella Rete;
 2. L'utente confermare l'operazione.

UC12: 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 preprocessare 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 agli 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.

UC13: Visualizzazione messaggio di errore per formato di dati non coerente con le aspettative

- **Descrizione:** Il file nel 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 nella non conformità del del formato di dati scelto;
- **Scenario principale:** Il file viene caricato all'interno della Rete.

UC4: Salvataggio della Rete

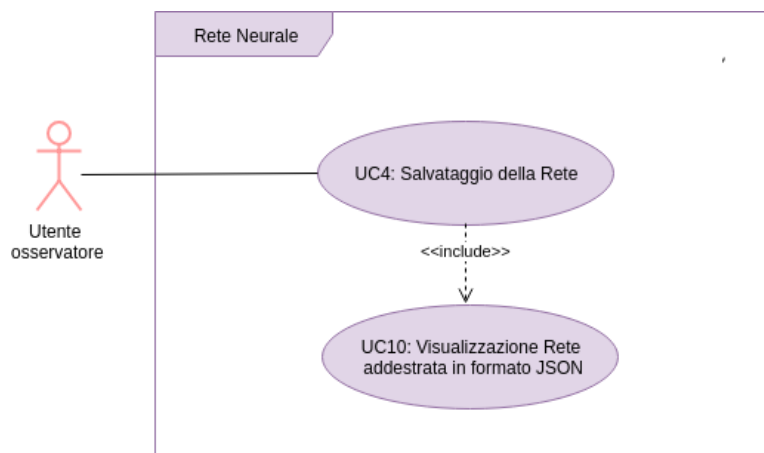


Figura 18: UC0_g: Operazioni utente - Visione generale interfacciamento con la Rete neurale (Salvataggio della Rete).

- **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:**
 1. L'utente effettua l'operazione di salvataggio della Rete.

UC10: 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:**
 1. L'utente visualizza i dati addestrati in formato JSON.

UC5: Caricamento della Rete

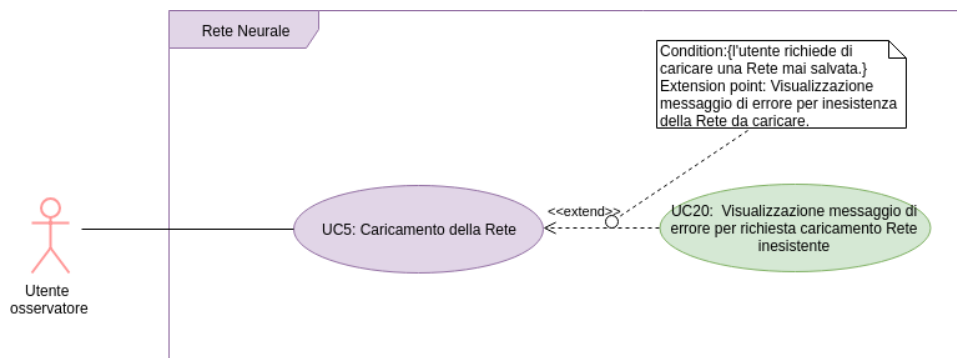


Figura 19: UC0_g: Operazioni utente - Visione generale interfacciamento con la Rete neurale (Caricamento della Rete).

- **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:**
 1. L'utente richiede l'operazione di caricamento della Rete precedente salvata.
- **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 della Rete da caricare richiesta.
- **Scenario principale:**
 1. All'utente viene indicato l'errore.

UC6: Visualizzazione batch di dati inseriti

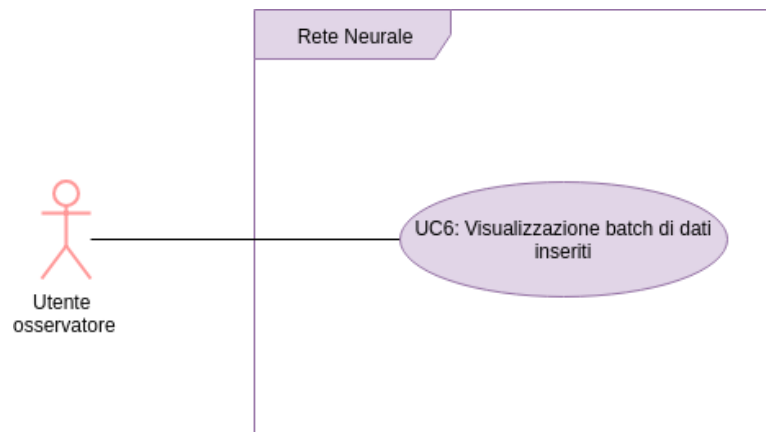


Figura 20: UC0_g: Operazioni utente - Visione generale interfacciamento con la Rete neurale (Visualizzazione batch di dati inseriti).

- **Descrizione:** L'utente visualizza i dati caricati nella Rete all'interno di una textarea riservata, all'interno della pagina web dell'applicativo; e 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:**
 1. L'utente visualizza i dati caricati nella Rete all'interno della textarea dell'applicativo della Rete 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 adempiere tale mansione. I dati caricati vengono presentati con indicazione del numero di test e di domanda a cui ognuno fa capo, cos  che all'utente risulti pi  agevole effettuare una veloce osservazione sugli stessi.
- **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:**
 1. L'utente visualizza i dati caricati nella Rete all'interno di una pagina web reindirizzata dall'applicativo della Rete stessa.

UC7: Modifica architettura della Rete

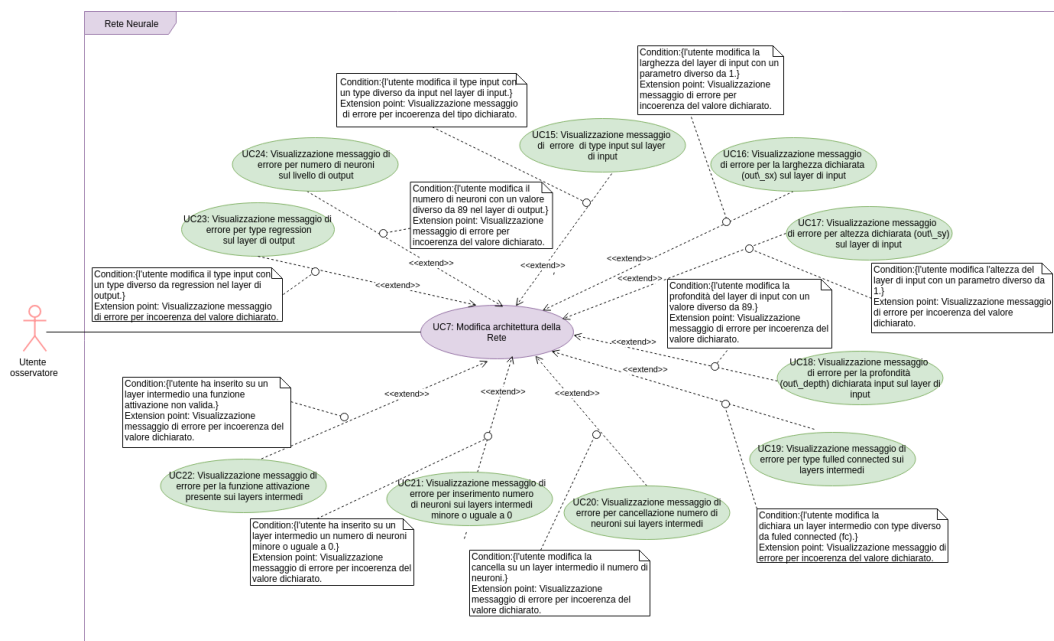


Figura 21: UC0_g: Operazioni utente - Visione generale interfacciamento con la Rete neurale (Modifica architettura della Rete).

- **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:** 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;
- **Scenario principale:**
 1. L'utente ha aggiunto un nuovo layer intermedio nella Rete (UC7.1);
 2. L'utente ha eliminato un layer intermedio esistente nella Rete (UC7.2);
 3. L'utente ha modificato un layer intermedio (UC7.3);
 4. L'utente ha modificato i parametri del trainer della Rete (UC7.4);
 5. L'utente conferma l'operazione di modifica della Rete (UC7.5);
- **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 (UC15);
 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 (UC16);
 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 (UC17);
 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 (UC18);
 5. L'utente ha tentato di modificare il tipo di connessione full connected su un layer intermedio della Rete. Viene ritornato un messaggio esplicativo che indica l'impossibilità di terminare l'operazione (UC19);
 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 (UC20);
 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 (UC21);
 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 (UC22);

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 (UC23);
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 (UC24).

UC7.1: 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:** 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 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;
 - (d) L'utente conferma l'operazione di modifica della Rete.

UC7.2: Eliminazione layer intermedio esistente

- **Descrizione:** L'utente ha la possibilità di riconfigurare l'architettura della Rete neurale. Esso può procedere a eliminare un layer intermedio presente nella 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 eliminando un layer intermedio;
- **Scenario principale:**
 1. L'utente ha eliminato un layer intermedio esistente nella Rete;

- (a) L'utente elimina il parametro di tipo;
- (b) L'utente elimina il numero di numero di neuroni;
- (c) L'utente elimina se presente la funzione attivazione;
- (d) L'utente conferma l'operazione di modifica della Rete.

UC7.3: Modifica layer intermedio esistente

- **Descrizione:** L'utente ha la possibilità di modificare un layer intermedio presente nella 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 provvedendo a ridefinire il contenuto di un layer intermedio;
- **Scenario principale:**
 1. L'utente ha ridefinito un layer intermedio presente nella Rete:
 - (a) L'utente può modificare il numero di neuroni presenti nella Rete;
 - (b) L'utente può modificare il parametro modificare/aggiungere/eliminare la funzione attivazione;
 - (c) L'utente conferma l'operazione di modifica della Rete.

UC7.4: 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 configurato 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;
- (e) L'utente conferma l'operazione di modifica della Rete.

UC15: 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 la incompatibilità del tipo richiesto con la natura della Rete.
- **Scenario principale:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC16: 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 la incompatibilità della dimensione richiesta con la natura della Rete.
- **Scenario principale:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC17: 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 la incompatibilità del della dimensione richiesta con la natura della Rete.
- **Scenario principale:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC18: 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 la incompatibilità del della dimensione richiesta con la natura della Rete.
- **Scenario principale:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC19: Visualizzazione messaggio di errore per type fulfilled 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 fulfilled 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 la incompatibilità del tipo richiesto con la natura della Rete.
- **Scenario principale:**

1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC20: 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 la incompatibilità dell'operazione richiesta con la natura della Rete.
- **Scenario principale:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC21: 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 la incompatibilità del valore definito con la natura della Rete.
- **Scenario principale:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC22: 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:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC23: 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:**
 1. All'utente viene indicato l'errore con l'indicazione della correzione necessaria per il corretto funzionamento dalla Rete.

UC24: 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:**
 1. 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, una volta 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 su cui effettuare con questi una 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).

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 (UC25);
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 (UC26);
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 (UC27);
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 (UC28).

UC9: Visualizzazione risultati della Rete

- **Descrizione:** L'utente visualizza i risultati ottenuti dalla Rete 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;
- **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 clusterizzazione 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).

UC9.4: 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 rappresentato 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 rappresentato 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.

UC11: Pulizia della textarea riservata a visualizzare lo storico dell'applicativo

- **Descrizione:** L'utente effettua la cancellazione dello storico in visualizzazione della 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:**

1. L'utente visualizza l'area di lavoro priva di informazioni sui dati elaborati fino a quel momento.

3 Principal Component Analysis

La **Principal Component Analysis** (acronimo PCA) è una tecnica impiegata nell'ambito della statistica multivariata² 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, che concorrono a comporre la *dimensionalità intrinseca*. 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 della variabilità presente nei dati rappresentabili in una singola dimensione, in poche parole 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, difatti conseguentemente 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 sui dati di trainset;

²parte 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 confronto della loro validità con le domande del database.

Ogni implementazione e osservazione l'ho svolta prima nei dati del trainset di prova e solo successivamente gli traslati all'interno del trainset delle domande nel database aziendale. In modo ho avuto sempre ben la correttezza e le 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 contenuti 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 sua 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 del trainset. Tale compito mi è risultato indispensabile perchè anche se di norma, la standardizzazione, viene usata per evitare situazioni erronee (alcune variabili X presentate con una variabilità molto maggiore rispetto ad altre) e io nei casi in analisi ero in possesso di dati già presentati con la medesima scala; avendo la necessità di individuare gli autovettori e applicare il criterio di Keiser nel trainset 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.

Metodologia che ho utilizzato per passi:

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 contenete per ogni variabile principale quali sono le componenti ordinate in modo decrescente strettamente correlate.

3.3 Risultati ottenuti

I dati risultanti dall'elaborazione dei dati inerenti il trainset delle domande nel database gli ho riportati, 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 0.0011520653 0.150101946 -0.693591199
V4 -0.3009414 -0.07357741 0.6112341 -0.0012775235 -0.719969817 -0.109796655
V5 0.3876316 0.53818077 0.2452635 -0.7069936980 -0.008478550 0.006089356
V6 -0.4905017 0.45668194 -0.2435273 0.0019475491 -0.152746837 0.684257823
```

Figura 22: Visualizzazione del calcolo della PCA tramite prcomp per il trainset di prova.

```
> prcomp(df_numeric, scale = FALSE)
Standard deviations (1, ..., p=89):
[1] 0.50804395 0.55848746 0.54971833 0.53126009 0.52009133 0.51577310 0.50144770 0.49349655 0.48596238 0.47999062 0.47018338 0.46401672 0.45887619 0.44999542 0.44252412 0.43580806 0.43223385 0.42680299
[19] 0.42079248 0.41533049 0.41068652 0.40526268 0.40119893 0.39676359 0.39238438 0.38808115 0.38384038 0.38044451 0.37616179 0.37446709 0.37095118 0.36811364 0.36109314 0.35803208 0.35352538 0.34981222 0.34587283
[37] 0.34221827 0.33861509 0.33532453 0.33275471 0.32925297 0.32449961 0.32227528 0.31788589 0.31716223 0.31403458 0.30962208 0.30523877 0.30243220 0.29816970 0.29458133 0.29251356 0.29083447 0.28424146
[55] 0.28177829 0.28074860 0.27888498 0.27585038 0.27422473 0.27139694 0.27019256 0.26679640 0.26427259 0.26314999 0.25905281 0.25708924 0.25674716 0.25387156 0.25253431 0.24954843 0.24591523 0.24355894
[73] 0.24107129 0.23709184 0.23498756 0.23280141 0.22959880 0.22384654 0.21341210 0.19833321 0.18012511 0.17449777 0.15236893 0.14016744 0.13282731 0.10140180 0.08865511 0.02703589 0.02334543

Rotation (n x k) = (89 x 89):
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10     PC11     PC12     PC13     PC14
V1 2.636387e-03 0.0053609165 -4.507406e-03 0.0047285077 -6.267813e-04 0.0012179541 -0.0006078271 0.0015559397 -1.333849e-03 0.0001061671 -0.0037520794 0.0133571060 -0.0124606511 0.016606655
V2 1.180602e-01 -0.0983273120 -2.06575e-01 -0.1595809648 1.666802e-01 -0.5078924212 -0.4156870156 -0.0055475540 2.878860e-01 -0.2664054093 -0.2141239846 -0.1154210418 -0.2867785700 -0.004759859
V3 2.418395e-03 -0.0075282117 1.23714e-02 -0.0130787491 3.664082e-03 -0.0173104077 0.0139035261 -0.0076250979 -1.464138e-02 -0.0268438815 -0.0315897238 0.0122222675 0.0332085902 -0.007904186
V4 -6.730258e-03 -0.0157451591 -2.170800e-02 -0.0117814649 4.12775e-02 0.0022788941 -0.0080299230 -0.0340036671 7.105564e-03 0.0206008584 -0.0081277662 0.0265778032 0.0022726268 -0.007387800
V5 1.514325e-02 -0.0057397594 -6.121892e-03 0.0045128218 5.433210e-03 0.0006548622 -0.0105138277 -0.0003915063 5.902137e-03 0.0027513825 0.0090012403 0.0390923132 0.0001094258 -0.007355540
V6 -3.397820e-02 -0.0335866990 1.107013e-02 -0.0932617778 2.604523e-05 0.0141769088 0.0067276127 -0.0041570833 -2.553817e-02 -0.0384170573 0.0118988680 -0.0180897858 0.0477093376 0.022818524
V7 1.314959e-02 -0.007094227 -3.653409e-02 0.0332109315 4.940004e-02 -0.0255328018 -0.0201294406 -0.040800858 7.001952e-02 -0.0055010060 -0.0186226671 -0.0534379835 -0.005495458 -0.009432314
V8 6.752564e-02 0.1271215300 -9.607290e-02 0.2670813592 2.783064e-02 -0.0864297064 0.0028403117 0.0012703572 -4.578910e-02 -0.0252448861 0.0348617417 -0.0236246302 -0.0258580808 -0.170862744
V9 4.916774e-02 0.0313629889 -6.180563e-02 0.1604337086 1.244642e-01 0.0159534435 0.0340172949 -0.0240708166 -7.438052e-02 -0.1016247255 -0.0391609350 -0.1892908371 0.1526616739 0.044115389
V10 1.203606e-02 -0.0068310303 -1.251339e-02 -0.0234524814 -2.780708e-02 -0.0023647650 -0.0193398945 -0.0257208548 3.119853e-02 0.0380631227 -0.0222724247 0.0529256183 -0.0389295281 -0.029166157
V11 -4.576335e-02 -0.0376339650 -7.111440e-02 -0.1541784131 -6.025191e-02 0.0093648752 -0.0358636247 -0.1687252230 -1.255892e-01 0.3334747609 -0.1520333807 -0.0680811778 -0.4090464042 -0.326270930
V12 1.707129e-02 0.1619149072 -1.751043e-01 -0.2636588899 -1.072001e-01 0.0471839798 0.0078876045 -0.0047098767 -1.602491e-01 -0.0972289384 0.1299995080 0.0078102762 0.0760172522 -0.007572510
V13 2.411796e-04 -0.0088153060 -1.048016e-02 -0.0395492127 -4.480856e-02 0.0199669461 0.0252357671 0.0631595225 8.353414e-03 -0.0116540106 0.0464408945 0.0549463204 0.0400919256 0.049566387
```

Figura 23: Visualizzazione del calcolo della PCA tramite prcomp per il trainset delle domande nel database.

Osservazioni Osservando esclusivamente la deviazione standard ottenuta dal trainset 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 trainset 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 24: Visualizzazione del calcolo del metodo summary ed individuazione degli autovalori per il trainset di prova.

```
> summary(res.pca)
Importance of components:
               Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6   Comp.7   Comp.8   Comp.9   Comp.10   Comp.11   Comp.12   Comp.13   Comp.14   Comp.15   Comp.16
Standard deviation   1.43136144 1.30933960 1.27248819 1.25540778 1.24609579 1.22900589 1.22206570 1.21814237 1.20907848 1.19278209 1.17999083 1.17045380 1.16024099 1.15735299 1.15660352 1.14908663
Proportion of Variance 0.02303868 0.01927807 0.01820818 0.01772265 0.01746071 0.01698505 0.01679377 0.01668611 0.01643872 0.01599857 0.0156336 0.01548521 0.01534701 0.01506696 0.01504486 0.01484788
Cumulative Proportion 0.02303868 0.04231675 0.06052493 0.07824758 0.09570828 0.11269334 0.12948710 0.14617321 0.16261193 0.17861059 0.1942441 0.20964931 0.22499632 0.24006328 0.25510814 0.26995602
               Comp.17   Comp.18   Comp.19   Comp.20   Comp.21   Comp.22   Comp.23   Comp.24   Comp.25   Comp.26   Comp.27   Comp.28   Comp.29   Comp.30   Comp.31   Comp.32
Standard deviation   1.13838196 1.12889068 1.12429911 1.11484336 1.10634870 1.10687053 1.09819751 1.09140282 1.0838406 1.07913609 1.07706163 1.07078478 1.06286985 1.05672587 1.05326482 1.05119497
Proportion of Variance 0.01457253 0.01433055 0.01421421 0.01397612 0.01376395 0.01375703 0.01356188 0.01339458 0.0132096 0.01309518 0.01304488 0.01289328 0.01270338 0.01255693 0.01247482 0.01242583
Cumulative Proportion 0.28452855 0.29885910 0.31307331 0.32704043 0.34081338 0.35457041 0.36813229 0.38152607 0.3947355 0.40783165 0.42087053 0.43376981 0.44647318 0.45903012 0.47150493 0.48393076
               Comp.33   Comp.34   Comp.35   Comp.36   Comp.37   Comp.38   Comp.39   Comp.40   Comp.41   Comp.42   Comp.43   Comp.44   Comp.45   Comp.46   Comp.47   Comp.48
Standard deviation   1.04333725 1.04116605 1.02871449 1.02479858 1.02307950 1.01728386 1.01558854 1.00733966 1.00662110 0.99744297 0.99650091 0.99578402 0.98376459 0.98205540 0.97349201 0.97125204
Proportion of Variance 0.01224076 0.01189987 0.01190005 0.01180962 0.01177003 0.01163706 0.01159812 0.01141066 0.01139439 0.01118755 0.01116643 0.01114858 0.01088282 0.01084503 0.01065672 0.01060774
Cumulative Proportion 0.49617152 0.50836139 0.52063144 0.53207106 0.54384509 0.55547815 0.56707628 0.57888604 0.58988133 0.60106088 0.61223531 0.62338389 0.63426671 0.64511174 0.65576687 0.66637621
               Comp.49   Comp.50   Comp.51   Comp.52   Comp.53   Comp.54   Comp.55   Comp.56   Comp.57   Comp.58   Comp.59   Comp.60   Comp.61   Comp.62   Comp.63
Standard deviation   0.96534807 0.95952978 0.95755890 0.95339737 0.94190469 0.93825159 0.93238471 0.93115086 0.92138027 0.91825097 0.91639097 0.90715173 0.90122820 0.89453214 0.89306972
Proportion of Variance 0.01047919 0.01032323 0.01031074 0.01022132 0.00997037 0.00999914 0.00977405 0.00974054 0.00974059 0.00954468 0.00948263 0.00944123 0.00923373 0.00913317 0.00898031 0.008968704
Cumulative Proportion 0.67685539 0.68728862 0.69751937 0.70774068 0.71771706 0.72761620 0.73739025 0.74714015 0.75668480 0.76616639 0.77560963 0.78486340 0.79399671 0.80299481 0.81196352
               Comp.64   Comp.65   Comp.66   Comp.67   Comp.68   Comp.69   Comp.70   Comp.71   Comp.72   Comp.73   Comp.74   Comp.75   Comp.76   Comp.77   Comp.78
Standard deviation   0.88780802 0.88008490 0.87705269 0.87186920 0.86541478 0.85970807 0.85674123 0.84783617 0.84675815 0.83708897 0.82968803 0.82365290 0.81733535 0.81241108 0.808770840
Proportion of Variance 0.00880612 0.00870979 0.00864984 0.00854744 0.00842181 0.00831147 0.00823383 0.00808319 0.00806264 0.00787754 0.00774889 0.00762046 0.00751209 0.00742182 0.007355447
Cumulative Proportion 0.82862816 0.82953615 0.83818649 0.84673442 0.85515629 0.86346740 0.87172132 0.87980451 0.88786764 0.89574678 0.90348756 0.91111622 0.91862801 0.92605012 0.93340572
               Comp.79   Comp.80   Comp.81   Comp.82   Comp.83   Comp.84   Comp.85   Comp.86   Comp.87   Comp.88   Comp.89
Standard deviation   0.79836916 0.79614608 0.78236719 0.77255190 0.75573703 0.75517304 0.74555504 0.73810154 0.71756860 0.69310962 0.644907608
Proportion of Variance 0.00716748 0.00712729 0.00688383 0.00671141 0.00649061 0.00641263 0.00621702 0.00596132 0.00579069 0.00540210 0.002430481
Cumulative Proportion 0.94057305 0.94770681 0.95458403 0.96129548 0.96776093 0.97419856 0.98041618 0.98637734 0.99216713 0.99756951 1.000000000
```

Figura 25: Visualizzazione del calcolo del metodo summary per il trainset 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 26: Individuazione degli autovalori per il trainset 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*. A tale scopo è prendere le componenti principali che catturano la maggior parte di variabilità dei dati. Nel caso del trainset di prova basta le variabili PC1, PC2 e PC3 sono sufficienti a catturare il 93% della variabilità presente. Quanto appena descritto si può riscontrare anche graficamente, come presentato di seguito.

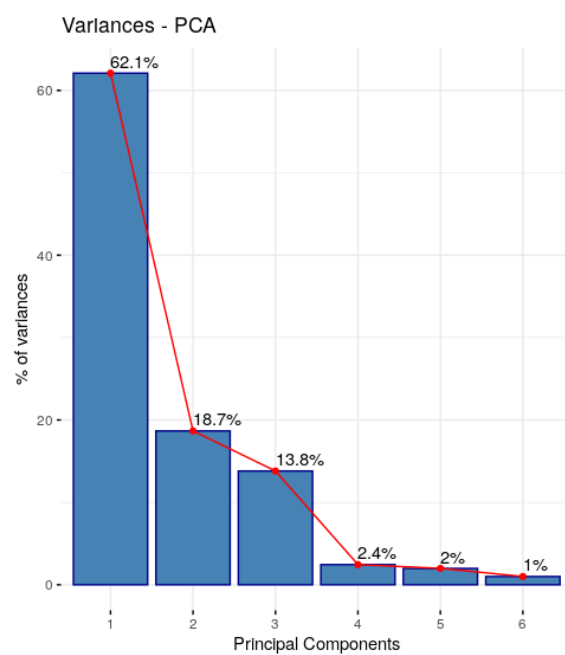


Figura 27: Plot della rappresentazione grafica della varianza sui PC del trainset di prova.

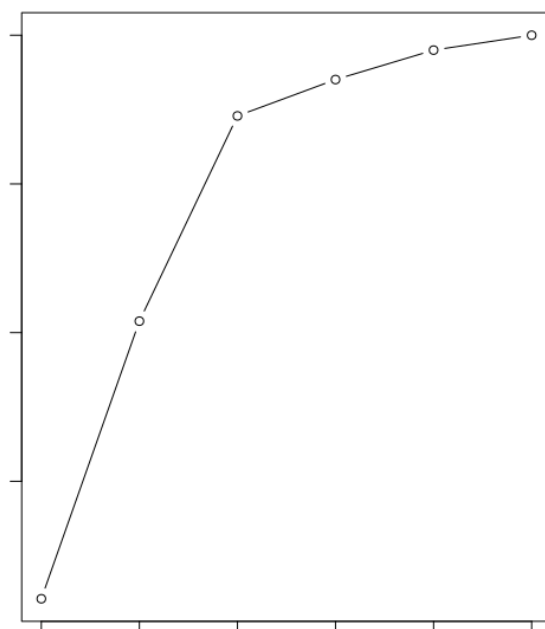


Figura 28: Plot della rappresentazione grafica della varianza sui PC del trainset di prova.

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

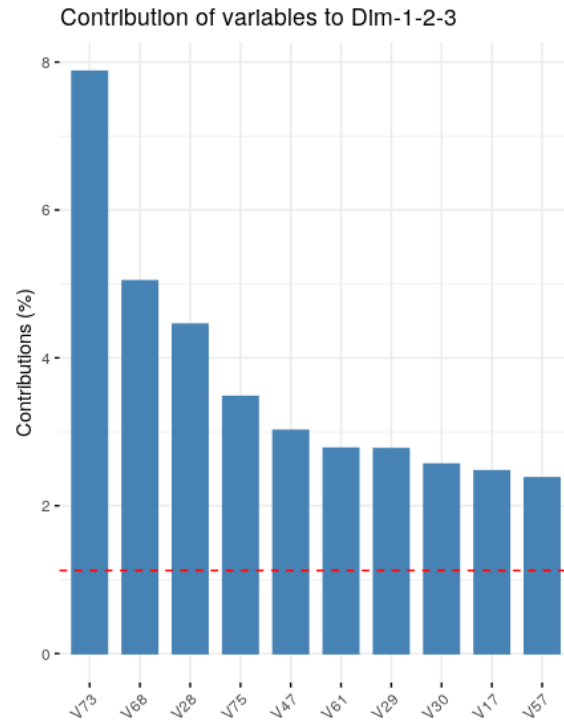


Figura 29: Rappresentazione grafica di come la varianza si distribuisce sulle PC individuate dal modello sul trainset di prova.

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

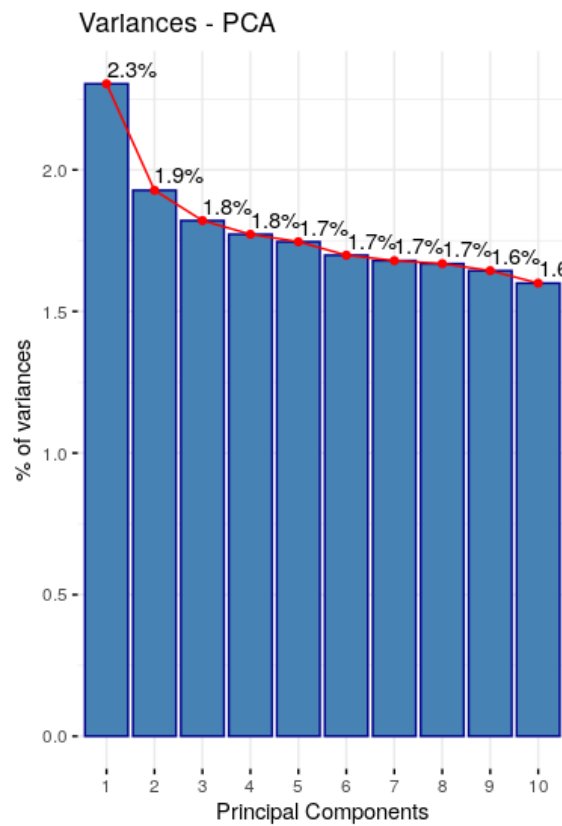


Figura 30: Plot della rappresentazione grafica della varianza dei primi dieci PC del trainset delle domande nel database.

Il plot mostra esclusivamente le prime dieci componenti; tutte si presentano con una varianza molto basse. 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.

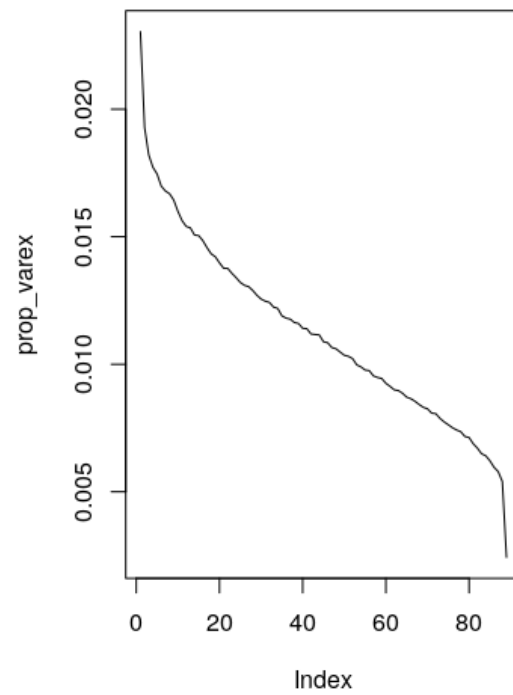


Figura 31: Plot della rappresentazione grafica della varianza di tutti i PC del trainset delle domande nel database.

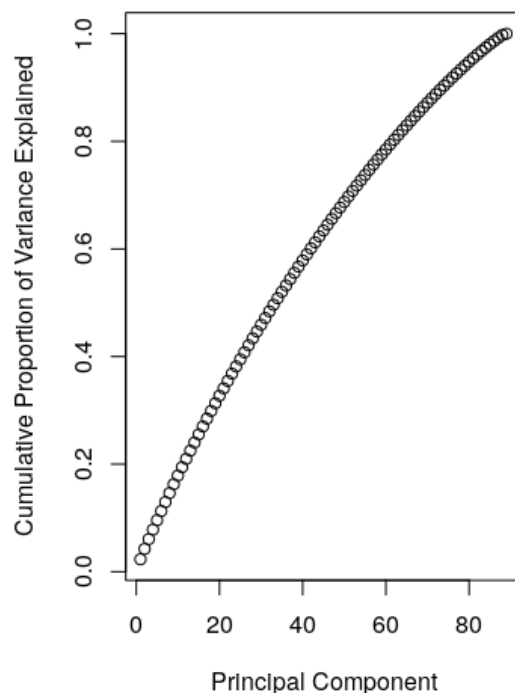


Figura 32: Plot della rappresentazione grafica della varianza di tutti i PC del trainset delle domande nel database.

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 effettuare una delle vie utilizzate è 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 stato in grado di dirmi molto, in quanto la diminuzione degli autovalori è graduale, senza salti evidenti. In conclusione, ho considerato come una percentuale di copertura adeguata quella fornita dalla che 41 prime componenti principali.

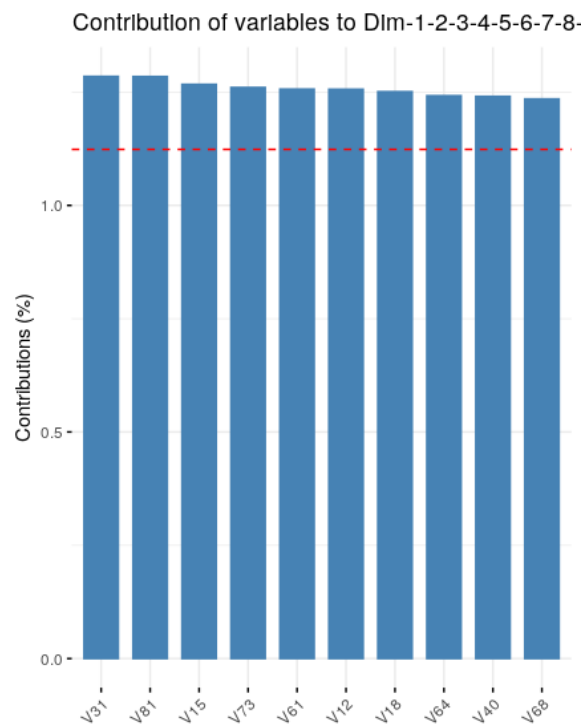


Figura 33: Rappresentazione grafica di come la varianza si distribuisce sulle PC individuate dal modello sul trainset delle domande nel database.

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 34: Autovettori del trainset 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 del trainset la prima variabile ha valori 0.420, 0.569 e 0.702, rispettivamente nelle componenti 1, 3 e 4 e sembra legarsi ai risultati nella 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 il trainset dei dati delle domande nel databas e per fare ciò ho provveduto a calcolare 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 35: Autovettori del trainset di test.

> cor_pca	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
V1	1.000000000	0.0208272106	-0.0065634941	-0.0090382409	0.03240232653	-1.222665e-02	1.440979e-02	0.0070758308	-0.06495159219	0.0096552727	0.0025737977	0.0066871180	-1.203887e-02	0.0061739747
V2	0.020827211	1.0000000000	0.0109844299	0.0277984172	0.0022019084	-1.355124e-02	0.014736e-02	0.0149978844	-0.0153876421	0.0310324650	-0.0327889285	-0.0105609866	-5.56337e-02	-0.0187739734
V3	-0.004563494	0.0109844299	1.0000000000	0.0099900228	0.00358048287	-5.024745e-03	-1.744317e-02	0.0096161144	0.0298033538	-0.0186844323	-0.0499445853	0.0118058675	4.296537e-03	-0.0123475985
V4	-0.009383294	0.0277984172	0.0099900228	1.0000000000	0.0070999842	-2.440739e-04	5.624869e-02	0.0022587458	0.0365871171	0.0152464032	0.0338313801	0.0355881260	-1.261759e-02	0.0213610839
V5	0.032402326	0.0022019084	0.0035804828	0.0070999842	1.0000000000	9.040432e-03	1.852404e-02	-0.0055583816	0.0257313911	-0.0075689377	-0.0020885953	0.0062431196	1.000551e-02	-0.0048544895
V6	-0.01222653	-0.013551263	-0.0058247451	-0.0002448739	0.0096045826	1.000000e-00	-9.60583e-03	-0.0244332634	0.0086920338	-0.0080880666	-0.0655248662	0.037934184	8.639492e-03	-0.0690937292
V7	0.014499783	0.0781473627	-0.0174431726	0.0562486863	0.0185240363	-9.605935e-03	1.000000e+00	0.0581811846	0.0185490010	0.0156306903	0.0231381746	-0.0031886528	-5.174636e-02	0.0362440823
V8	0.007075831	0.0149978844	0.0096161144	-0.0022587458	-0.0055583816	-2.443236e-02	5.618110e-02	1.0000000000	0.076588911	0.0078533344	0.0256721371	-0.0263441964	-2.351978e-02	-0.0004265940
V9	-0.064951592	-0.0153876421	0.0298033538	0.0365871171	0.0075689377	0.014736e-02	0.0149978844	0.076588911	1.0000000000	-0.0349939777	-0.0142172292	-0.0101187252	-2.265859e-02	-0.0084913600
V10	0.009653273	0.0310324650	-0.0186844323	-0.0075689377	-0.0088667e-03	1.563069e-02	0.0078533144	-0.0349939777	-0.0349939777	1.0000000000	0.0527671384	0.0174195214	-1.752897e-02	-0.0315645566
V11	0.002557398	-0.0327889285	-0.0499445853	0.0318330681	-0.0020885953	-6.552441e-02	2.118117e-02	0.0256721371	-0.0142172292	0.0527671384	1.0000000000	0.0086376189	-9.989418e-02	0.0476110206
V12	0.00687118	-0.0105609866	0.0118058675	0.0355881260	0.0062431196	3.779342e-02	-3.108653e-03	-0.0263441964	-0.0101187252	0.0174195214	0.0086376189	1.0000000000	-2.087564e-02	-0.0436117988
V13	-0.0123881878	-0.0555637880	0.0042965374	-0.0126175820	0.0100855066	8.639492e-03	-5.174636e-02	-0.0235197807	-0.0226585934	-0.0175289660	-0.0998940992	-0.0287553534	1.000000e+00	-0.0118653285
V14	0.0061739775	-0.0187739734	-0.0123475985	0.0213610839	-0.0048544895	-6.909373e-02	3.624695e-02	-0.0084265948	-0.0084913600	-0.0315645566	0.0476118206	-0.0436317988	-1.186524e-02	1.0000000000
V15	-0.074558925	-0.0108939765	0.02641120927	-0.0093846234	-0.0238124172	-3.009299e-02	1.114677e-02	0.0207760267	-0.0392981147	-0.0222128800	-0.0017877176	0.0361480356	-6.254951e-02	-0.0093746618
V16	-0.02662358	-0.011545265	0.0565738729	-0.0469258098	0.0042767939	1.902876e-03	1.274478e-03	0.0114727247	-0.0178050337	-0.0376931682	-0.0025752982	0.0019073041	-4.278461e-02	-0.0735133872
V17	-0.022553887	0.019581396	0.0232952095	0.0151158566	0.0135916823	-6.883897e-02	-1.720176e-02	0.0732774112	0.0398040762	-0.0153210876	0.0053478215	-0.0043380998	-9.066885e-02	0.0243573872
V18	0.033595719	0.015947767	0.0154470300	-0.0282878561	0.0568743229	-3.990590e-03	4.486876e-02	-0.0312988401	0.0308294586	-0.0086338392	0.0232946240	-0.0251171795	7.200024e-03	0.0003392265
V19	0.07722819	0.0616471697	0.0205788512	-0.0260978417	0.0235088060	1.955257e-02	5.402097e-03	-0.0094760426	-0.0013380927	0.0165682352	-0.0138078684	0.0263225378	6.487972e-02	0.0178439408
V20	-0.104791476	0.0498297021	-0.0035088826	-0.006636388	-0.144212569	-8.77787e-02	-5.282266e-02	0.0051956413	-0.0043515419	0.0070749883	0.0063787773	-0.427326e-02	0.0045376852	
V21	0.001301325	0.0316287927	-0.0186170659	0.0240706152	-0.0010221703	-1.603287e-02	6.974488e-02	0.0514436061	0.0165919084	-0.0139277357	0.0504173164	0.0490217371	-5.489776e-03	0.0508444488
V22	0.045962996	-0.0423980226	0.0268650232	-0.0098374767	0.0006265081	-4.037566e-02	7.071899e-03	0.0407826922	-0.0136334834	-0.0072728183	-0.0258236358	-0.0311185464	-3.022208e-02	-0.0307062098
V23	0.009565787	-0.013444095	0.0444239593	-0.0161160728	-0.0075857749	5.642315e-02	-4.119466e-03	0.0143374178	-0.0122236390	-0.0277880811	-0.0205181241	-0.0428572731	-3.804852e-02	-0.0450139111
V24	0.020861554	0.0025415277	-0.0078510543	0.0115072959	0.0140898602	-5.18751e-02	1.604833e-02	0.0522678520	0.0180121917	0.0098716182	-0.0209595968	-0.0309977718	-1.424688e-03	0.0386945664
V25	0.010275598	-0.0351132590	-0.0423664687	-0.0102464149	0.0061111783	-4.992188e-02	2.756454e-02	0.0375284283	-0.0062977412	-0.0208749214	-0.0208813582	0.0080121162	-1.886551e-02	-0.0064415288
V26	-0.00976081	0.0437080951	-0.0204118478	-0.0195245796	0.0050682468	-2.641208e-02	-3.814904e-02	0.0091286432	-0.0847883208	-0.0022588353	0.0172217957	-0.0024255320	-3.725737e-02	-0.0108715181
V27	0.001958346	0.0052576995	-0.0289586575	-0.0389506607	-0.0051383680	-3.752493e-02	-4.667907e-02	0.0657728135	-0.0641613837	0.01919144912	0.0206141402	-0.0158782337	-0.1958761e-02	0.0192883889
V28	0.060859885	0.0056544438	0.0016281298	0.0032246160	0.0842716139	4.362135e-03	1.617656e-02	-0.0025244628	0.0231730686	-0.0034376016	-0.0009124095	-0.00389993260	4.580558e-03	-0.0022847745
V29	-0.006088069	0.000480793	-0.0050801786	-0.0069323137	0.0242400271	-0.0046541414	-0.0046541414	-0.00578642298	-0.0754163816	-0.0107088482	-0.0096558480	-0.0008423465	-1.426399e-02	-0.0068683339
V30	0.081173322	0.024601493	0.0007919649	0.0100453533	-0.0582349537	1.158896e-02	-4.936794e-02	0.0078642298	-0.0754163816	-0.0107088482	-0.0096558480	-0.0008423465	-1.426399e-02	-0.0068683339
V31	0.015643566	-0.018521426	-0.0145959643	-0.0217811370	-0.0112063393	-9.158019e-02	1.525326e-02	-0.0111650752	0.0325952484	-0.0395465069	0.0292287789	-0.0238433864	-5.295257e-02	0.0140466496
V32	0.024680801	-0.0038273079	-0.0189354505	-0.0380029342	-0.0072842951	-3.782271e-02	2.223626e-02	0.0735098278	0.0488627662	-0.0496173499	0.0028080101	0.0206264840	-4.796123e-03	0.01350462879
V33	0.004874525	-0.0149725427	0.0338943818	-0.0465115468	-0.0032007214	-4.091823e-02	6.711966e-02	0.0508707391	-0.0647437596	-0.0299316827	0.0313914808	-0.0001254868	-2.493290e-02	0.0401322686
V34	0.016266451	-0.0446870482	-0.002401855	0.0027819968	-0.0130688211	-1.589349e-02	1.116960e-02	-0.0984235780	-0.0174180527	-0.0393713876	0.0083573898	0.0045888559	-3.054788e-02	-0.0469246356
V35	-0.040844583	-0.0506293448	-0.0097860238	0.0067496286	0.0282693454	1.843719e-02	-6.455991e-02	-0.044088329	-0.0469380450	-0.0450905899	-0.0265444517	-0.0325695266	-4.813517e-02	-0.0173156735
V36	-0.010432892	-0.0765002870	-0.0024631210	-0.017285893	0.0082041441	-3.007722e-02	1.356876e-02	0.0077480409	-0.0267110404	0.0514032384	-0.0553808467	0.0331457637	-1.799453e-02	-0.0192302794
V37	0.009568084	-0.0150858025	0.0635131458	-0.0176091642	-0.0075811384	-2.354418e-02	-4.576751e-02	0.0035853161	0.0228089750	-0.0492936022	0.0062671749	-0.029875684	-2.218056e-02	-0.055363615
V38	-0.009846157	0.0236583883	-0.0359366022	0.0281873376	0.0238927681	1.929486e-02	-1.481809e-02	-0.0224116215	0.0585327987	0.0066000548	-0.0129748364	0.0131885226	6.800904e-03	-0.0242852181
V39	0.012155212	-0.029519701	-0.0027640867	-0.0548243741	-0.0095400355	9.961818e-03	1.438521e-03	-0.0217956515	-0.0191110988	-0.0348691871	0.0028995906	-0.0039131474	-0.012021e-03	0.0153979974
V40	0.049559619	-0.027333676	-0.017460880	-0.0206436799	-0.0346388446	-2.305803e-03	-3.861638e-02	-0.0453786369	-0.0372820895	-0.0288445639	-0.0526078138	-0.022386315	6.927169e-02	-0.0273261579
V41	0.011096954	-0.0374918888	-0.0134970214	-0.0374110271	-0.0087171539	1.582526e-03	8.971930e-03	-0.0360435448	-0.0271766879	0.0253216642	0.0275827018	-0.006476182	-2.287136e-02	-0.026616937

Figura 36: Autovettori del trainset delle domande nel database.

I CSV generati di correlazione sono i seguenti:

	V1	V2	V3	V4	V5	V6
1	1	4	3	6	5	2
2	2	5	6	3	1	4
3	3	6	1	4	3	6
4	4	3	6	1	4	3
5	5	6	3	1	4	3
6	6	3	1	4	3	6

Figura 37: CSV generato a partire dalla matrice correlazione del trainset della rete di prova.

I file CSV gli ho generati prendendo la matrice correlazione, che mostra il grado di correlazione di ogni componente con ogni variabile del modello, creandovi 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).

Da tale distribuzione dei dati, per il trainset 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 del trainset del database, tuttavia questo non è stato possibile, a causa delle seguenti motivazioni:

- 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 nel database) tale fenomeno verrebbe espanso impedendo 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 del trainset di prova soddisfano appieno le attese (infatti non solo le correlazioni matchano perfettamente con quanto viene indicato dal grafo della conoscenza; ma anche il plot che viene generato dalla PCA sulle variabili presenta una coerenza stretta con tali assunzioni) 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 il trainset di prova generato da n input sottoposti alla probabilità di indovinare. Come viene illustrato dai plots nella sezione seguente, i dati individuati dal trainset 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 triplice 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.

Infatti 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.

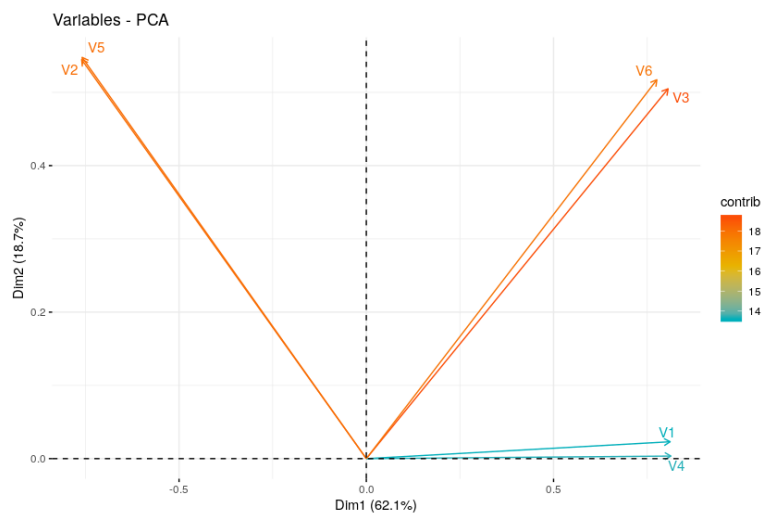


Figura 39: Rappresentazione per mezzo di plot di come le variabili si presentano nelle due componenti principali con il calcolo della PCA - utilizzo di trainset di prova puro sul grafo della conoscenza.

```
> 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 40: Le domande in correlazione con la variabili principali in ordine decrescente sui dati di prova.

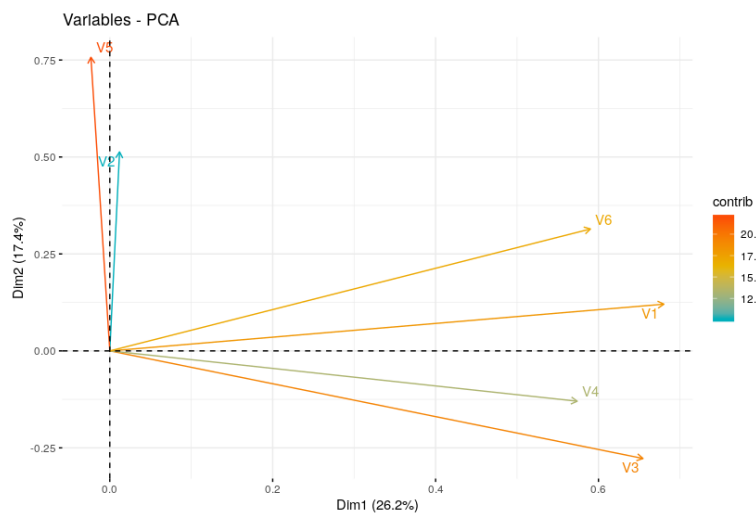


Figura 41: Rappresentazione per mezzo di plot di come le variabili si presentano nelle due componenti principali con il calcolo della PCA - utilizzo di trainset di prova spurio con la probabilità di indovinare.

```
> 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 42: Le domande in correlazione con la 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 e studiare i cluster possibili per mezzo delle previsioni assunte da ogni variabile sulla base della previsione standard³.

³Vettore previsione tutto a zero

4 Costruzione del Reticolo della Conoscenza

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

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 questa prima riga che viene interpretata come una dichiarazione e conseguentemente non rappresentata all'interno del Reticolo.

Nel sistema possono venire settati i seguenti aspetti:

- La *tipologie* 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 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 permette all'utente di decidere se si desidera procedere con una rappresentazione del Reticolo manuale o automatica progressiva dei dati.

4.1.1 Configurazione

Analizzando l'applicativo in base al carattere dei dati in ingresso e alle aspettative sull'output del modello, ho riscontrato che la configurazione necessaria per la formazione del Reticolo della Conoscenza è vincolata alla dichiarazione delle seguenti proprietà:

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

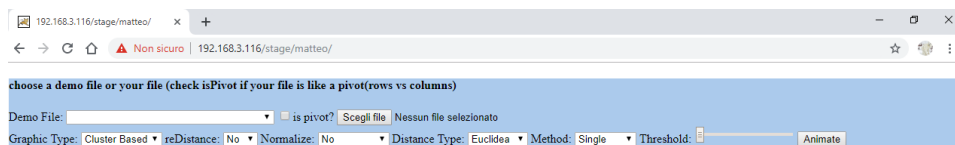


Figura 43: Configurazione usata nel sistema per generare il Reticolo della Conoscenza.

5 Creazione dei file CSV

Come già accennato all'interno della sezione §4.1 prima di procedere alla creazione del Reticolo ho dovuto preparare i dati di previsione prima di poterli dare in pasto al sistema. 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 di standard ⁴.

Ho fatto in modo che il vettore delle differenze venga stampato su console del browser ⁵, 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.

6 Creazione del Reticolo della Conoscenza per sui dati di Prova

A	B	C	D	E	F	G	H	I	J	K	L	M
1	1371737561165000	0.308506184988141	0.0718524387566007	0.1389571230802850	0.3126620179583976	0.08758773376353950	0.1253914756230576	0.2819224585022953	0.0454503930947081	0.1269235506446909	0.285727283023205	0.08156486500918674
2	2284480752632270	0.4804501730256432	0.1821449074859530	0.2365583540634790	0.481398676466212	0.15271287586977390	0.2683952418226050	0.5835950645662583	0.1315988961708890	0.275968608451425	0.576248188579408	0.1805638000523014
3	0.1598860302661289	0.11114630022487592	0.8835251671568856	0.15118474206351700	0.1237286795832643	0.6456771916893269	0.1005178086725840	0.1305539493313499	0.5070337812452888	0.0936656242975840	0.1407066413661771	0.47900142657547656
4	0.2340515577799740	0.40902108186467346	0.480351693909651	0.2310208064321070	0.25818230961250716	0.435017616105854	0.198149522282257	0.23312599756181320	0.36778641640621130	0.1938403571014482	0.231743364899585	0.3471598910572275
5	0.0071117007597130	0.489923446691022	0.3951556170266950	0.0762238077876230	0.5045525781400051	0.37380554109711490	0.12077474818167620	0.6837887139900635	0.31414112599189020	0.195261210548219	0.15886903026496674	0.297225393020335
6	0.2282250939155690	0.02848928657081700	0.8097741507867975	0.21879540139615590	0.041580814887648890	0.748433103009213	0.1518574810388590	0.04657932412718770	0.5740459136881284	0.14495193550482490	0.056340675956571290	0.5422182071530703
7												

Figura 44: file CSV generato per la creazione del Reticolo della Conoscenza sui dati di prova.

⁴Vettore tutto a zero

⁵unica alternativa facendo uso di solo codice javascript

TODO: SCREEN DEI VARI PASSAGGI DI FORMAZIONE DEL RETICOLO PER LA RETE DI PROVA

6.0.1 Osservazioni

6.0.2 Osservazioni Reticolo dati di prova

Il Reticolo della Conoscenza è perfettamente coerente con le aspettative; ricalca fedelmente quanto evidenziato nella figura 6 nella sezione § 2.1 Inoltre gli accoppiamenti tra i punti ricalcano quanto dichiarato dalla matrice correlazione ottenuta dal calcolo del modello effettuato con la PCA e visibile dall'immagine 37. Tuttavia i risultati della PCA non devo essere presi come assioma in quanto la Rete neurale, può cogliere oscillazioni che il modello matematico non è in grado.

7 Creazione del Reticolo della Conoscenza sui dati delle domande nel database

	A	B	C	D	E	F	G	H	I	J	K
1											
2	beta3wzq	0.0288012759232788	0.0326033964119093	0.07836415966691804	0.006657077138899262	0.014203260806503760	0.014538420752104130	0.00041633948122965500	0.016482964308414453	0.01675811290371937	0.03101155005072145
3	beta2wzq	0.0207181189664192	0.15814753027738458	0.0725849498604126	0.04945749734785253	0.03791026584094914	0.006257452941515240	0.0216407787597845	0.031224581343231875	0.0150227050209037	0.0329279170895644
4	beta3wzq	0.0208802234034442	0.0670379077688657	0.112441608417004	0.0424377773278654	0.013602512117038014	0.01077032956460200	0.0305418189951173	0.03309747318031809	0.0139445254643046	0.031827020886773
5	beta3wzq	0.00462733047028937	0.0517977544018809	0.041291371255641440	0.07590169862822324	0.01324984784970831	0.0354174496745808	0.016841260991757164	0.015784832634415173	0.047155959787122	0.0106735855651816
6	beta3wzq	0.00530145729121291	0.02360144131718002	0.00954696230247950	0.0150209547403931	0.03744045433085234	0.0120541514880953	0.021971441863953686	0.014323047285762017	0.0353314729428718490	0.0109974487456464
7	beta3wzq	0.000402057920868120	0.005952180847914383	0.009547376112080470	0.03246031260802036	0.004562674390578990	0.0291897949493174	0.013324674800112081	0.013274822949528448	0.033908506574687628	0.001732599957522389
8	beta3wzq	0.0362093201354507	0.000939357901115483	0.0230580170272695	0.0103584763049027	0.0416850765316308	0.01320496047867480	0.12706074817561195	0.0104540472448135	0.05481734966618888	0.0240751477579714
9	beta3wzq	0.01397205472462978	0.02610074379242598	0.001583717298378840	0.017491208786054	0.0102045253213053	0.00473996954430880	0.0160475646698926	0.16313646453547884	0.054706550393864	0.00740394540691225
10	beta3wzq	0.02097149901403808	0.04217140369391018	0.05660021824151509	0.01290623851225042	0.0073905546971465150	0.00475546969621090	0.015602356752599	0.018205215469434093	0.0075007173443040108	0.0202138658070484
11	beta3wzq	0.03012533931319097	0.12625485140887965	0.0225754464310478320	0.07823142641725508	0.00588707723014029	0.013489900853270630	0.04650889433624325	0.014169662336526999	0.00879920964531622	0.00496531380701643
12	beta3wzq	0.00441930445038545	0.12345281571988028	0.0989699017424209	0.04044643442151289	0.02529175279570940	0.01758635006768880	0.00862271024190057	0.00819722525976738	0.01521277459454102	0.0158035190243428
13	beta3wzq	0.0127641334632812	0.02494448179730562	0.0311991807602547	0.0178802317713105	0.00210925849000140	0.01239930897769919	0.08655025454684	0.04611544733329296	0.03842173493202474	0.0108189404079529
14	beta3wzq	0.019719256672752914	0.0530254310324654	0.03524633254264182	0.025363834980308773	0.05230428203833045	0.00312076825010496	0.04901149848464743	0.005646699705159526	0.04623849474004114	0.01702937897347977
15	beta3wzq	0.00023640425400807	0.02001404265819723	0.0180772432208874	0.00025416134980337	0.01006114007780785	0.00519358986240084	0.0470340496549511	0.00921418189251461	0.02644796700847797	0.000274545773817576
16	beta3wzq	0.00035329275880555	0.04505299251888664	0.01262137474071321	0.018736509501428	0.029735701986269362	0.012138677788661190	0.01885132460460002	0.02837278173213198	0.01856215464026823	0.00839611704964199
17	beta3wzq	0.01015742320835	0.030945230839961	0.0443823106299005	0.033379146588778	0.00732338171300140	0.005014725768496	0.0655436930897583	0.02472086476029404	0.00415652115154843	0.0056535630312559
18	beta3wzq	0.0001357514952578	0.007837118489547366	0.0410306078930551	0.05229128602395109	0.013836107174227398	0.0317742801255134	0.01769199878626758	0.0506065971752995	0.03423075112828001	0.01167243217987458
19	beta3wzq	0.0135377758364377	0.01119235270311204	0.0210609359245536	0.0173265389311968	0.001078756726239290	0.01053296700684790	0.029712502859545817	0.002142702417622965	0.0115214180418932868	0.0037412634242776
20	beta3wzq	0.01806145197408013	0.032086095868175	0.0401539113218189	0.00330446869213030	0.01762452539787781	0.0069951777948696	0.0412174772370074	0.01447778126800727	0.017084303041112	0.00378241913858508
21	beta3wzq	0.01847200776342973	0.03609085927225	0.0194570054732277	0.02882698054018142	0.019389128711565173	0.02175409373974478	0.07321548163081411	0.04570069051309324	0.050252471668897	0.0070776764131452
22	beta3wzq	0.0425324112084578	0.06044858303151386	0.014359116185229290	0.02499011082484196	0.0004139767869808	0.02209710972695228	0.139705529286918	0.013796980615841563	0.0717781606705208	0.0655635630312559
23	beta3wzq	0.00028855069800600	0.0023144831928867	0.00563452161890960	0.034745985705629	0.0050924308109428290	0.030593884797956130	0.02258755815042811	0.017007182848575116	0.0417643602962854	0.00340800525910704
24	beta3wzq	0.02216798735601	0.04361341532970286	0.040742071019161	0.0168408552949097	0.00934084298488800	0.01204236252397890	0.0308010633981313	0.01104986443639936	0.0268070307195479	0.00314517831469929
25	beta3wzq	0.0062415474469893	0.00846410879138962	0.0237551184662440	0.046207876974451	0.01237184810696384	0.02968878414257405	0.046645737778619266	0.00032247486024825380	0.0448471370401046	0.0139931015364666
26	beta3wzq	0.002168440271063480	0.000337349115197165	0.000337349115197165	0.01317051044249054	0.015541902909690	0.0000529653110516250	0.03439896828712683	0.003984200618953634	0.0098409193407105	0.0109101666085007
27	beta3wzq	0.00310910823140981	0.041162436077363	0.00364113613650180	0.0477468024363405	0.00240268139963384	0.00670067026251140	0.032723699101024	0.04122307017542525	0.009454045007664	0.0032779145509039
28	beta3wzq	0.00068870879367277	0.0384402275950796	0.0678627131363057	0.02186481059883250	0.0044574010954830790	0.01299898560471600	0.00813281137062025	0.014122580589508984	0.02443297327769574	0.0216162297325932
29	beta3wzq	0.043017479890952	0.039955442259873	0.0777013087152079	0.01504263157605197	0.0381945780740481	0.021486079733225290	0.0693999270081418	0.02099661511433415	0.01395818603745148	0.0109313469809468
30	beta3wzq	0.00069893277804	0.0029477554553023	0.0103416810697114	0.0261786829222342	0.01535974168481810	0.004002759122052	0.03996134390597804	0.0170199579476888	0.0548823071670763	0.01761290472049
31	beta3wzq	0.0119885457414349	0.0234837968412975	0.041013567707129	0.0389917812257722	0.0337180302389236	0.0330575501953240	0.0452806199654182	0.0149027115502031319	0.010102735696028	0.00988030702629408
32	beta3wzq	0.0109917953771413	0.01507790450310762	0.042900438884895	0.053022118405791	0.068138425738358	0.033991481625727790	0.01181344487031731	0.02852140734578963	0.0489532550469074	0.00333107951450099
33	beta3wzq	0.00509215340023735	0.0176763688808533	0.01648710715750876	0.00911885150682297	0.01084492463200807	0.00355422054423820	0.0086193144405915	0.0580869193407105	0.00113566621047105	0.0109101666085007
34	beta3wzq	0.0086847929007878	0.0006627815315973	0.01338428700720078	0.023343328028697	0.02381071547154126	0.011991723138642816	0.0083233047626884	0.059432174803077966	0.0388952594933445	0.0060805058909017
35	beta3wzq	0.0102621444403975	0.06203503548778736	0.04970742618432356	0.013831849902388931	0.03931657521065124	0.001602213989613910	0.016021669625500348	0.06303748472934512	0.00019550909706823	0.025013226488824
36	beta3wzq	0.0102669113075	0.0374634169517773	0.0211256691113075	0.00234507377715240	0.017802016572576	0.01420948028346230	0.02529194493958195	0.014711804411063102	0.00013595909706823	0.025013226488824
37	beta3wzq	0.0108014999741567	0.05278167661690159	0.01878195323878429	0.01844315646033869	0.00024498734206980	0.00743401017160140	0.00573362543217613	0.0495257989389644	0.00972378625018881	0.0054099877290183
38	beta3wzq	0.00370047355549416	0.004480733432739012	0.004712317526992490	0.009378732044678	0.001987690331320798	0.002889092256676290	0.028461471090497126	0.010562873082452169	0.005640461401378777	0.00329655112244287

Figura 45: Porzione di esempio file CSV generato per la creazione del Reticolo della Conoscenza sui dati del database.

La creazione del Reticolo della Conoscenza sui dati delle domande nel database ha richiesto la creazione di file CSV su un architettura a 2 layer della rete con 6, 8, 10, 12 neuroni ciascuno. Questo a causa della non conoscenza a priori del numero di cluster che comporranno il Reticolo, come invece accade per i dati di prova.

Tale differenziazione ha permesso durante la configurazione del sistema e la creazione del Reticolo che mi accorgersi dell'estrema variabilità dei cluster e correlazioni a seconda dell'oscillazione del numero di neuroni per layers, come viene evidenziato dalle immagini seguenti.

TODO: INSERIMENTO DEI SCREEN DEL RETICOLO SUI CASI A 6, 8, 10, 12 NEURONI A CLUSTER E A FORCE BASED.

7.1 Osservazioni

7.1.1 Osservazioni Reticolo dati del database

La variazione dei Reticoli ottenuti dalle diverse architetture della Rete neurale indica come sia necessario approfondire la tematica e individuare una strategia unica che permetta di individuare in modo univoco l'architettura necessaria per ottenere un Reticolo della Conoscenza valido.