

Text_Sepsis_SIRS_EDITED

File computazione 2024-05-21 test1.R

Numero campioni 1257

Dimensioni 16

Media, mediana, min, max

Calcola la media per tutte le colonne numeriche

```
media_totale <- colMeans(load_Text_Sepsis_SIRS_EDITED)
```

Calcola la mediana per tutte le colonne numeriche

```
mediana_totale <- apply(load_Text_Sepsis_SIRS_EDITED, 2, median)
```

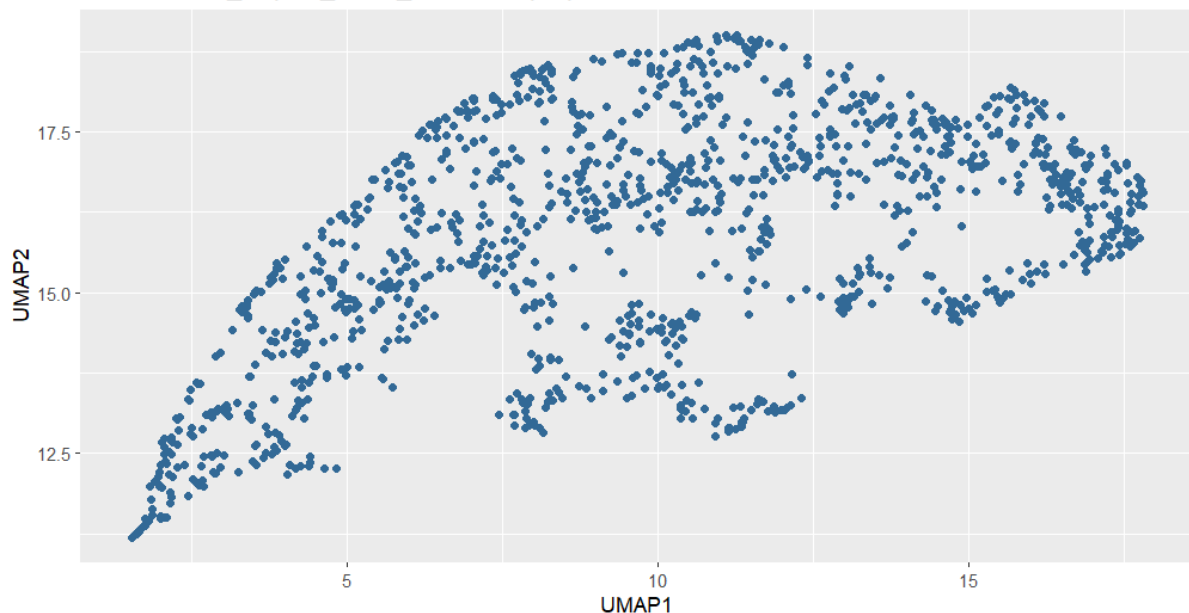
Calcola il valore minimo e massimo per tutte le colonne numeriche

```
minimo_totale <- apply(load_Text_Sepsis_SIRS_EDITED, 2, min)
```

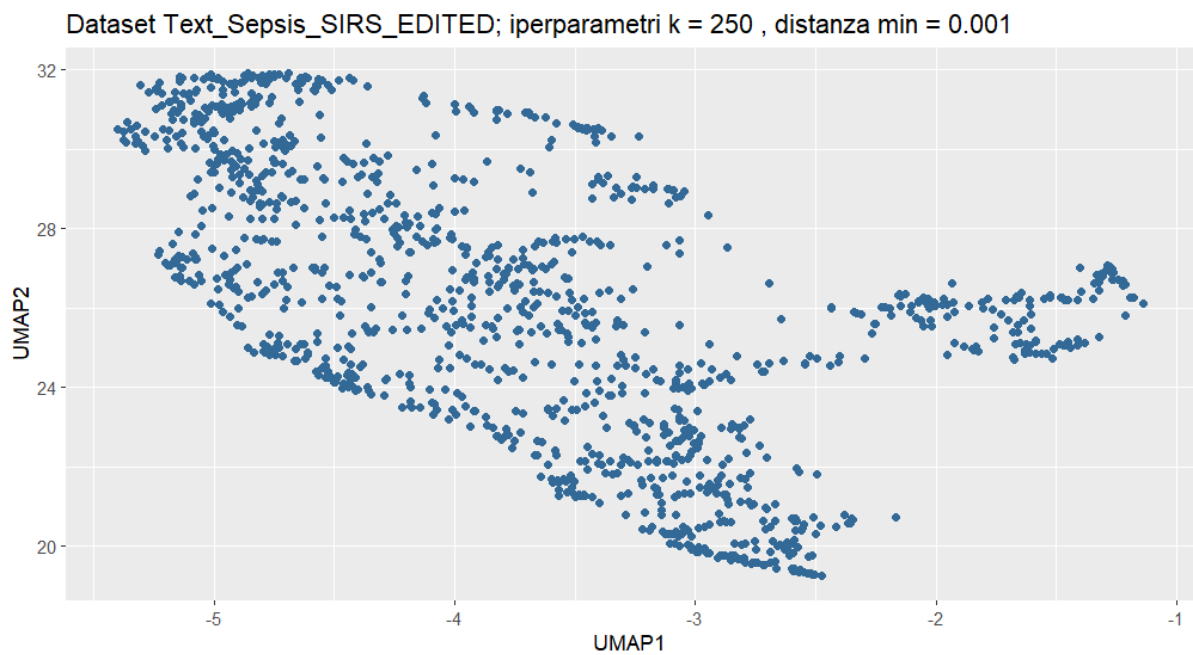
```
massimo_totale <- apply(load_Text_Sepsis_SIRS_EDITED, 2, max)
```

Iperparametri Struttura locale

Dataset Text_Sepsis_SIRS_EDITED; iperparametri k = 30 , distanza min = 0.2

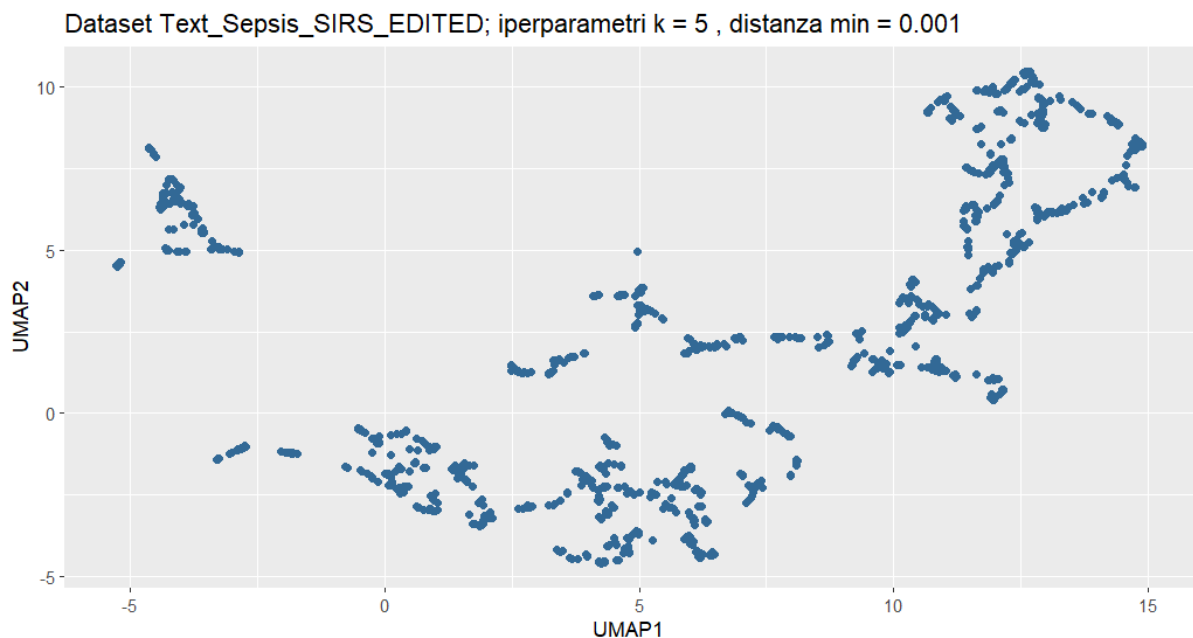


Iperparametri Struttura globale



Con questi valori possiamo vedere che una piccola parte dei punti forma un cluster sulla destra

Iperparametri Ottimo



Si possono vedere 4 clusters.

Individuo i clusters

Adesso individuo con una valutazione visiva, delle coordinate che mi permettano manualmente di separare i punti in questi 4 clusters.

Ricordo che le proiezioni UMAP cambiano ad ogni run, quindi i punti si spostano e potrebbero non coincidere con le coordinate scelte per la suddivisione.

cluster1

```
mio_grafico + geom_point(x = 0, y = 2.5, color = "red", size = 3)
```

cluster2

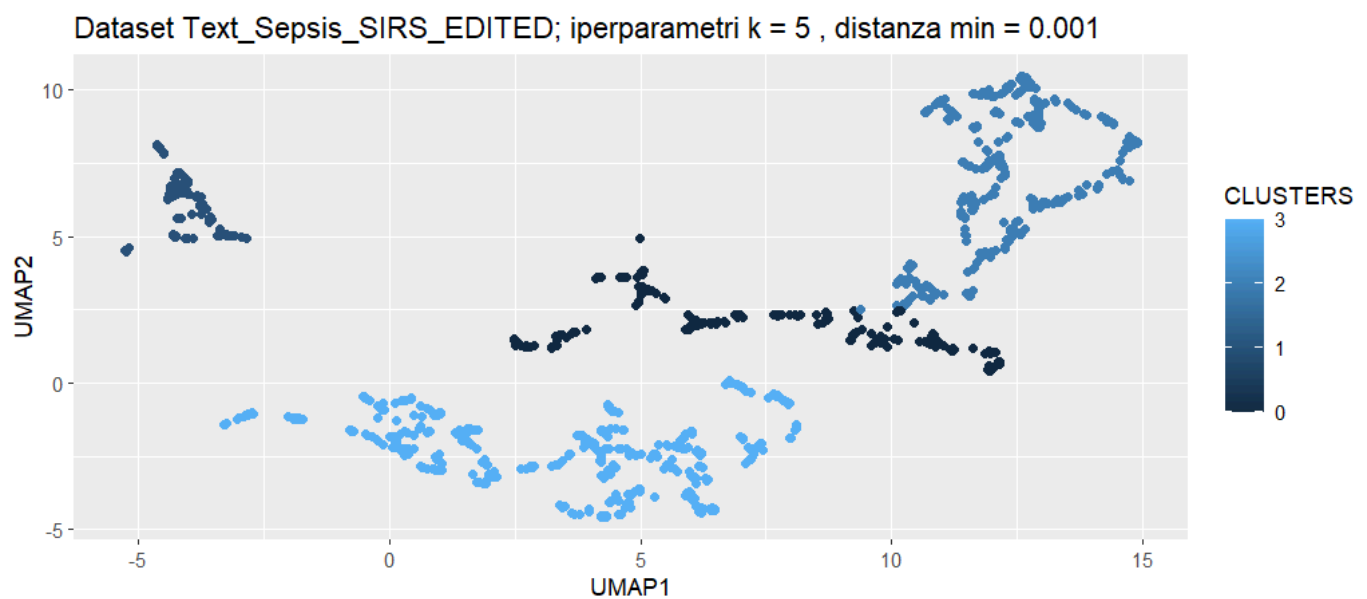
```
mio_grafico + geom_point(x = 9, y = 2.5, color = "red", size = 3)
```

cluster3

```
mio_grafico + geom_point(x = 9.7, y = 0.4, color = "red", size = 3)
```

cluster0

Sono i rimanenti punti.



statistiche per i clusters

```
### Seleziono il cluster 1
```

```
### merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 1,]
```

media

```
media_cluster0 <- colMeans(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 0,])
```

```
media_cluster1 <- colMeans(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 1,])
```

```
media_cluster2 <- colMeans(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 2,])
```

```
media_cluster3 <- colMeans(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 3,])
```

```
statistiche <- as.data.frame(media_totale)
```

```
statistiche <- cbind(statistiche, as.data.frame(media_cluster0))
```

```
statistiche <- cbind(statistiche, as.data.frame(media_cluster1))
```

```
statistiche <- cbind(statistiche, as.data.frame(media_cluster2))
```

```
statistiche <- cbind(statistiche, as.data.frame(media_cluster3))
```

	media_totale	media_cluster0	media_cluster1	media_cluster2	media_cluster3
Age	56.5059666	61.7384615	59.21487603	57.1875000	52.38655462
sex_woman	0.4081146	0.3884615	0.48760331	0.4100000	0.39705882
diagnosis_OEC_1M_2_AC	0.5465394	0.5961538	0.75206612	0.4950000	0.51050420
APACHE II	13.3365155	14.4115385	14.57851240	14.8675000	11.14705882
SOFA	2.5823389	2.8076923	2.23140496	3.5725000	1.71638655
CRP	6.1850835	6.8227692	8.22776860	5.6959000	5.72859244
WBCC	11.7679419	11.6593462	15.16520661	9.3148750	13.02506933
NeuC	9.9004972	9.7531154	12.51471074	7.9532500	10.95280462
LymC	1.0345195	1.0535769	1.51608264	0.7363625	1.15224790
EOC	27.2394590	59.5000000	40.49586777	18.4500000	13.63445378
NLCR	13.1166269	12.3857692	12.38842975	15.1742500	11.97184874
PLTC	194.3619730	178.0500000	422.71900826	88.2825000	234.36554622
MPV	10.1277407	10.3165385	9.92644628	10.2868250	9.94210084
Group	0.3508353	0.3153846	0.46280992	0.4325000	0.27310924
LOS-ICU	4.4359586	4.1846154	6.08264463	4.9650000	3.71008403
Mortality	0.1026253	0.1076923	0.09917355	0.1450000	0.06512605

La differenza maggiore tra i clusters sembra essere sulle dimensioni PLTC, LOS-ICU, EOC, WBCC.

minimo

```
minimo_cluster0 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 0,], 2, min)
```

```
minimo_cluster1 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 1,], 2, min)
```

```
minimo_cluster2 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 2,], 2, min)
```

```
minimo_cluster3 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 3,], 2, min)
```

```
statistiche <- as.data.frame(minimo_totale)
```

```
statistiche <- cbind(statistiche, as.data.frame(minimo_cluster0))
```

```
statistiche <- cbind(statistiche, as.data.frame(minimo_cluster1))
```

```
statistiche <- cbind(statistiche, as.data.frame(minimo_cluster2))
```

```
statistiche <- cbind(statistiche, as.data.frame(minimo_cluster3))
```

	minimo_totale	minimo_cluster0	minimo_cluster1	minimo_cluster2	minimo_cluster3
Age	18.000	18.00	19.00	19.000	18.00
sex_woman	0.000	0.00	0.00	0.000	0.00
diagnosis_0EC_1M_2_AC	0.000	0.00	0.00	0.000	0.00
APACHE II	0.000	1.00	1.00	2.000	0.00
SOFA	0.000	0.00	0.00	0.000	0.00
CRP	0.000	0.00	0.05	0.020	0.01
WBCC	0.600	0.75	3.09	0.600	2.49
NeuC	0.200	0.35	2.09	0.200	1.75
LymC	0.005	0.10	0.15	0.005	0.12
EOC	0.000	0.00	0.00	0.000	0.00
NLCR	0.500	1.10	0.50	0.900	1.00
PLTC	11.000	55.00	312.00	11.000	133.00
MPV	0.000	0.00	8.00	0.000	7.70
Group	0.000	0.00	0.00	0.000	0.00
LOS-ICU	1.000	1.00	1.00	1.000	1.00
Mortality	0.000	0.00	0.00	0.000	0.00

La differenza maggiore tra i clusters sembra essere sulle dimensioni PLTC, WBCC, NeuC, LymC.

massimo

```
massimo_cluster0 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 0,], 2, max)
```

```
massimo_cluster1 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 1,], 2, max)
```

```
massimo_cluster2 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS == 2,], 2, max)
```

```
massimo_cluster3 <- apply(merged_df[, 1:(ncol(merged_df) - 3)][merged_df$CLUSTERS ==
3,], 2, max)
```

```
statistiche <- as.data.frame(massimo_totale)
```

```
statistiche <- cbind(statistiche, as.data.frame(massimo_cluster0))
```

```
statistiche <- cbind(statistiche, as.data.frame(massimo_cluster1))
```

```
statistiche <- cbind(statistiche, as.data.frame(massimo_cluster2))
```

```
statistiche <- cbind(statistiche, as.data.frame(massimo_cluster3))
```

	massimo_totale	massimo_cluster0	massimo_cluster1	massimo_cluster2	massimo_cluster3
Age	99.00	98.00	96.00	99.00	98.00
sex_woman	1.00	1.00	1.00	1.00	1.00
diagnosis_OEC_1M_2_AC	2.00	2.00	2.00	2.00	2.00
APACHE II	48.00	46.00	36.00	48.00	39.00
SOFA	16.00	12.00	10.00	16.00	13.00
CRP	52.05	47.46	51.50	50.53	52.05
WBCC	51.08	37.36	46.32	51.08	45.59
NeuC	50.50	35.41	39.59	50.50	40.56
LymC	6.97	6.97	5.80	5.54	5.29
EOC	410.00	410.00	230.00	140.00	80.00
NLCR	420.80	74.90	111.50	420.80	55.20
PLTC	854.00	348.00	854.00	221.00	342.00
MPV	107.00	13.40	80.20	107.00	12.70
Group	1.00	1.00	1.00	1.00	1.00
LOS-ICU	96.00	55.00	48.00	96.00	84.00
Mortality	1.00	1.00	1.00	1.00	1.00

La differenza maggiore tra i clusters sembra essere sulle dimensioni PLTC, WBCC, NeuC, EOC, NLCR, MPV, LOS-ICU.