

project model 3 model based clustering and model comparison

Jingjing Li

2022-12-16

I have trouble plot figures on model density, uncertainty, classification The error message is “figure margin too large” I tried the following methods: `dev.off()` `par(mar=c(1,1,1,1))` drag the plot area bigger I also tried just load a part of the dataset however, these method did not work

Helper packages

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.7      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(stringr)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##      combine
```

```
library(cluster)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(mclust)
```

```
## Package 'mclust' version 6.0.0
## Type 'citation("mclust")' for citing this R package in publications.
##
## Attaching package: 'mclust'
```

```
##
## The following object is masked from 'package:purrr':
##
##      map
```

process the data

```
df <-read.csv("radiomics_completedata.csv")

df <- na.omit(df)

df<-select(df,-c(Institution, Failure.binary))

set.seed(123)
```

model based clustering

```
mdf<- Mclust(df, G = 3)
a<-summary(mdf) #if print a, too many pages

plot(mdf, what = "density")
```

```
## Error in plot.new(): figure margins too large
```

```
plot(mdf, what = "uncertainty")
```

```
## Error in plot.new(): figure margins too large
```

```
sort(mdf$uncertainty, decreasing = TRUE) %>% head()
```

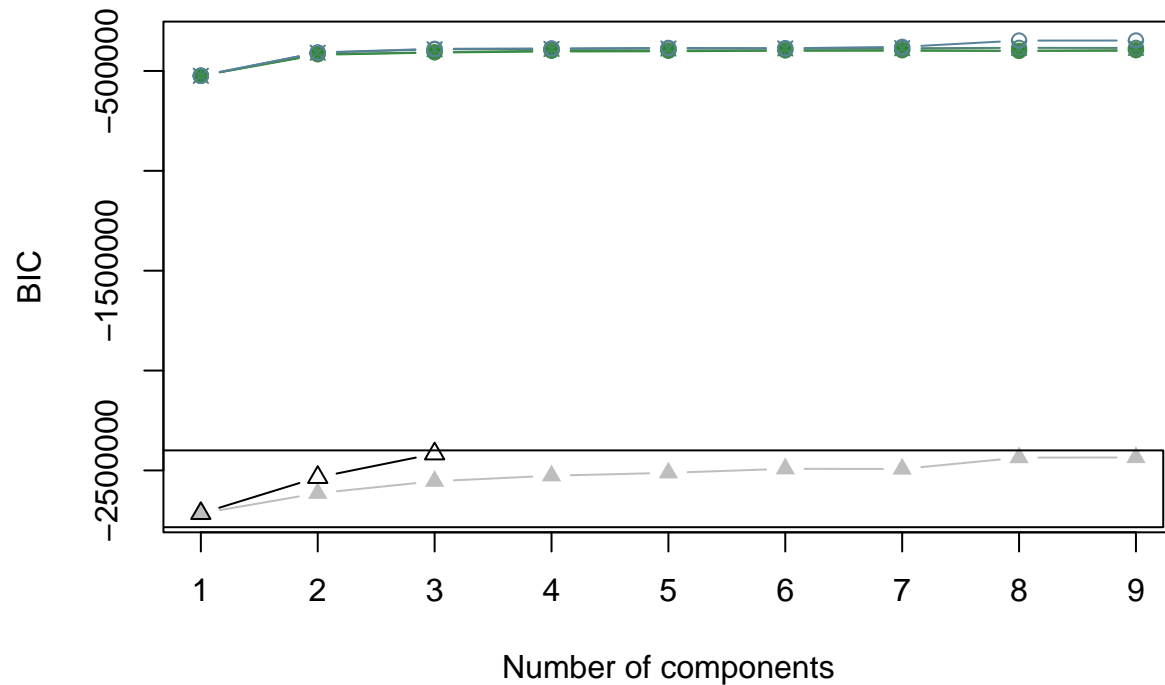
```
##           100           93           48           4           1           2
## 3.485515e-07 5.757084e-11 5.693224e-13 8.881784e-16 0.000000e+00 0.000000e+00
```

```
mdf_mc <- Mclust(df)

summary(mdf_mc)
```

```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVI (diagonal, varying volume and shape) model with 9 components:
##
## log-likelihood  n  df      BIC      ICL
##    -153331.3 197 7730 -347501.7 -347501.7
##
## Clustering table:
##  1  2  3  4  5  6  7  8  9
## 94 38  3  2 10  8 15  9 18
```

```
legend_args <- list(x = "bottomright", ncol = 427)
plot(mdf_mc, what = 'BIC', legendArgs = legend_args)
```



```
plot(mdf_mc, what = 'classification')
```

```
## Error in plot.new(): figure margins too large
```

```
plot(mdf_mc, what = 'uncertainty')
```

```
## Error in plot.new(): figure margins too large
```

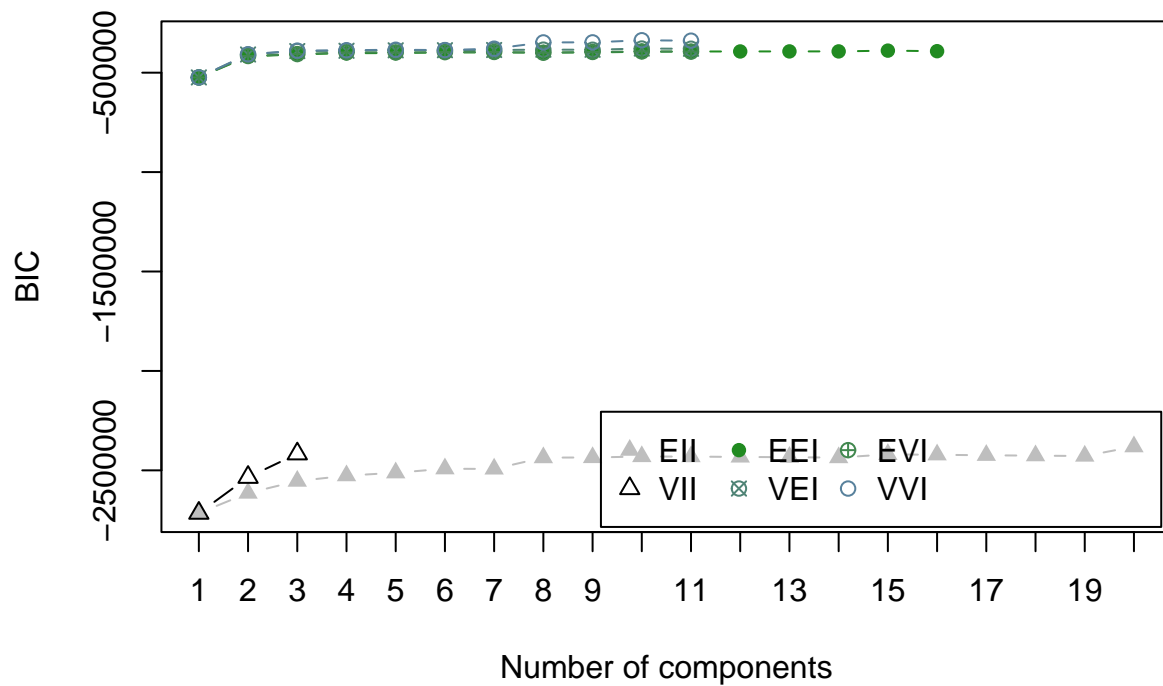
```
df_mc <- Mclust(df, 1:20)
```

```
summary(df_mc)
```

```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVI (diagonal, varying volume and shape) model with 10 components:
##
## log-likelihood  n  df      BIC      ICL
##      -145818.9 197 8589 -337015.1 -337015.1
```

```
##
## Clustering table:
## 1 2 3 4 5 6 7 8 9 10
## 56 39 37 3 2 10 8 15 9 18
```

```
plot(df_mc, what = 'BIC',
      legendArgs = list(x = "bottomright", ncol = 5))
```

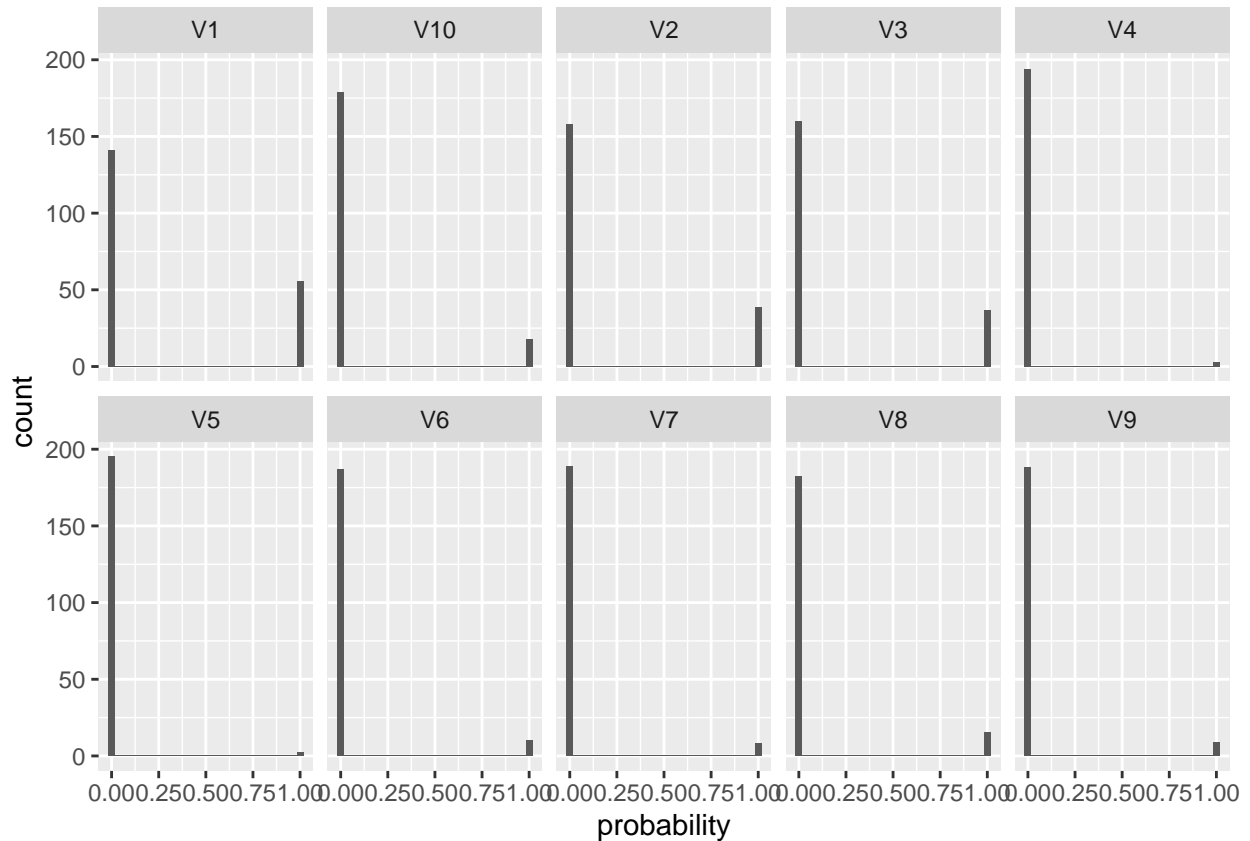


```
probabilities <- df_mc$z

probabilities <- probabilities %>%
  as.data.frame() %>%
  mutate(id = row_number()) %>%
  tidyr::gather(cluster, probability, -id)

ggplot(probabilities, aes(probability)) +
  geom_histogram() +
  facet_wrap(~ cluster, nrow = 2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
uncertainty <- data.frame(
  id = 1:nrow(df),
  cluster = df_mc$classification,
  uncertainty = df_mc$uncertainty
)

cluster2 <- df %>%
  scale() %>%
  as.data.frame() %>%
  mutate(cluster = df_mc$classification) %>%
  filter(cluster == 2) %>%
  select(-cluster)

cluster2 %>%
  tidyr::gather(product, std_count) %>%
  group_by(product) %>%
  summarize(avg = mean(std_count)) %>%
  ggplot(aes(avg, reorder(product, avg))) +
  geom_point() +
  labs(x = "Average standardized consumption", y = NULL)
```

