Brain clustering and probability estimation

Chanwoo Lee, September 24, 2020

1 Clustering method

From the coefficient $\boldsymbol{B}$, I perform clustering based on SVD. Since the output $\boldsymbol{B}$ is symmetric (I made $\boldsymbol{B}$ symmetric from outputs $\boldsymbol{P}_{\text{row}}$ and $\boldsymbol{P}_{\text{col}}$). The coefficient $\boldsymbol{B}$ has the

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^T, \text{ where } \boldsymbol{P} \in \mathbb{R}^{d\times r} \text{ is an orthonormal matrix and } \boldsymbol{\Sigma} \in \mathbb{R}^{r\times r} \text{ is a diagonal matrix.}$$

I perform K-means clustering on scaled principle components $\boldsymbol{P}' = \boldsymbol{P}\sqrt{\boldsymbol{\Sigma}}$. Notice that similar rows of $\boldsymbol{P}'$ ensure the similar rows of $\boldsymbol{B}$, implying the corresponding nodes having similar impact on response variables.

2 Brain clustering: IQ brain data

Based on the previous result, I set the (rank, cost) = (8,1), (I mistook to set rank 8 instead of 3 or 4 so I requested new job setting the rank 3) calculate the coefficient $\boldsymbol{B}$, and perform K-means clustering. The following figure is the elbow plot on $\boldsymbol{P}'$. The elbow method suggests 6 clusters among the 68 nodes.

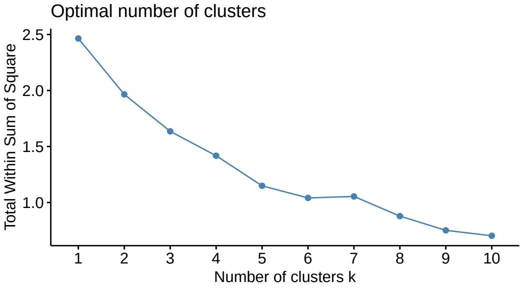


Figure 1: The elbow plot for determining the number of clusters in K-means.

Based on Figure 1, I set the number of clusters as 4,5,6 and check the PCA results and brain image results for each case. Figure 2 shows brain nodes according to top two principle vectorsn and Figure 3 plots the brain nodes according to the clusters.

It seems to me that when the number of clusters is four, we can see clear separation between the left and right hemispheres. As the number of clusters increases, the separation is a little bit vague.

3 Brain clustering: VSPLOT brain data

For this dataset, setting rank is a little bit vague to me so I choose 4 different ranks according to a different criterion. First, I set the rank 8 at which hinge loss cross validation graph shows local minimum. Second, I set the rank 30 at which 0-1 loss shows local minimum. Third, I set the rank 49 at which 0-1 loss has the global minimum. Lastly, I set the rank 68 which is the full rank. To summarize, the elbow plots cannot suggest the number of clusters because it decreases with the

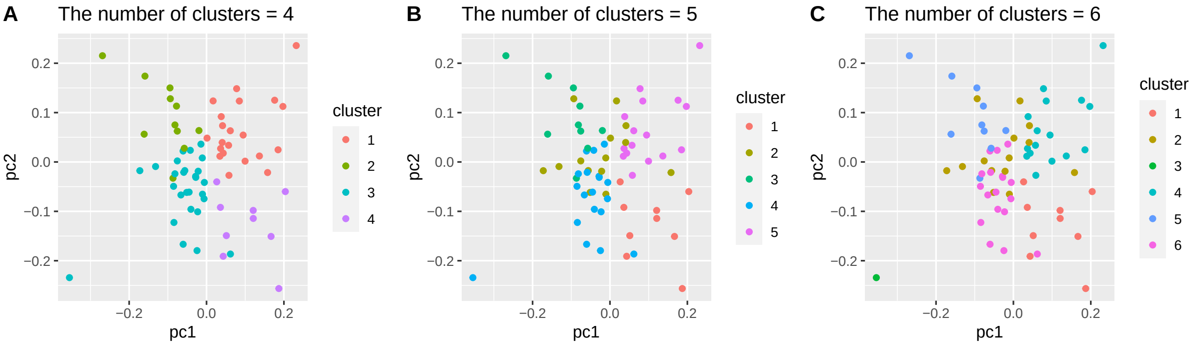
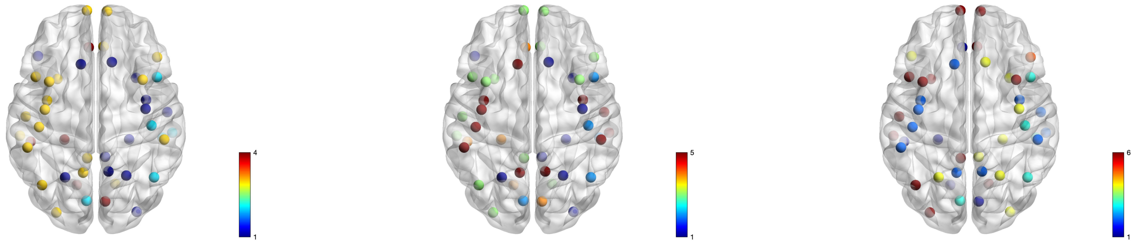


Figure 2: Clustering results according the number of clusters. The first axis is the largest principle vector and the second on is the second largest principle vector. A: when the number of clusters = 4, B: when the number of clusters = 5, and C: when the number of clusters = 6.



(a) The number of clusters = 4 (b) The number of clusters = 5 (c) The number of clusters = 6

Figure 3: Figure of brain nodes according to clusters.

same ratio regardless of the group size. Therefore, I clustered brain nodes into 4 to 5 groups and checked the brain regions. It seems to me that when the group size 4, our clustering shows the good separation of brain regions except the full rank case.

3.1 Clustering results when rank = 8

The following is the elbow plot when rank = 8. There is slightly flat part around when group size is around 5 and around 8. So I check the clusters when the group size is 4,5, and 6. When the group size is 7 or 8 are in Appendix.

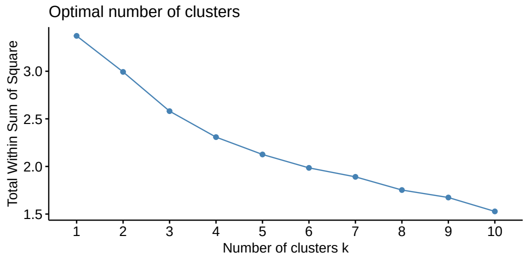


Figure 4: The elbow plot for determining the number of clusters in K-means.

I check the PCA results and brain image results for each case. Figure 5 shows brain nodes according to top two principle vectors and Figure 6 plots the brain nodes according to the clusters.

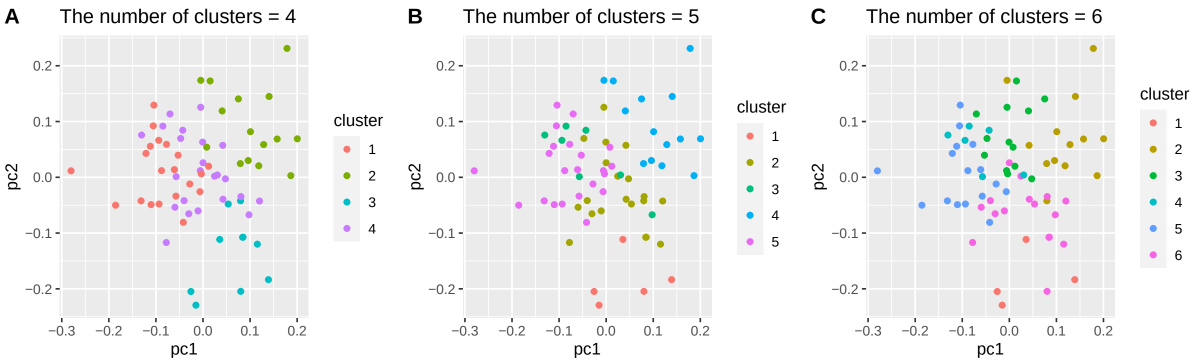
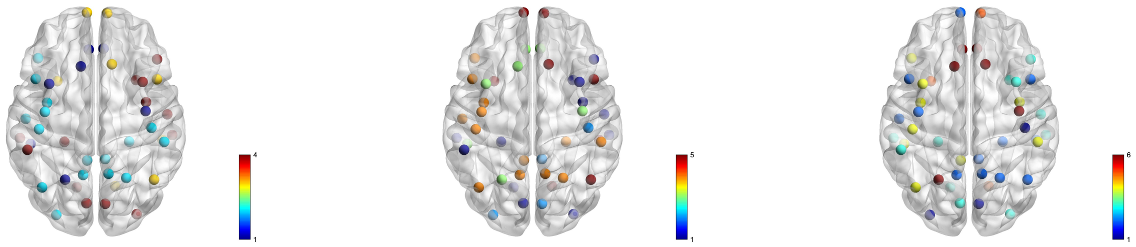


Figure 5: Clustering results according the number of clusters. The first axis is the largest principle vector and the second on is the second largest principle vector. A: when the number of clusters = 4, B: when the number of clusters = 5, and C: when the number of clusters = 6.



(a) The number of clusters = 4 (b) The number of clusters = 5 (c) The number of clusters = 6

Figure 6: Figure of brain nodes according to clusters.

3.2 Clustering results when rank = 30

The following is the elbow plot when rank = 30. The elbow plot does not give us the optimal group size because it decreases consistently with same rate according to the group size. So I stick to the group size 4,5, and 6 as before. When the group size is 7 or 8 are in Appendix.

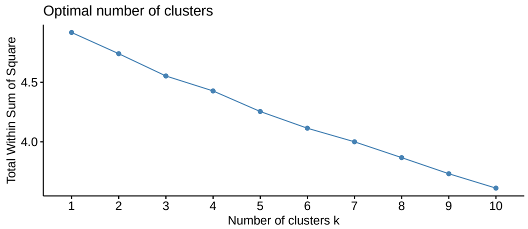


Figure 7: The elbow plot for determining the number of clusters in K-means when rank = 8.

I check the PCA results and brain image results for each case. Figure 8 shows brain nodes according to top two principle vectors and Figure 9 plots the brain nodes according to the clusters.

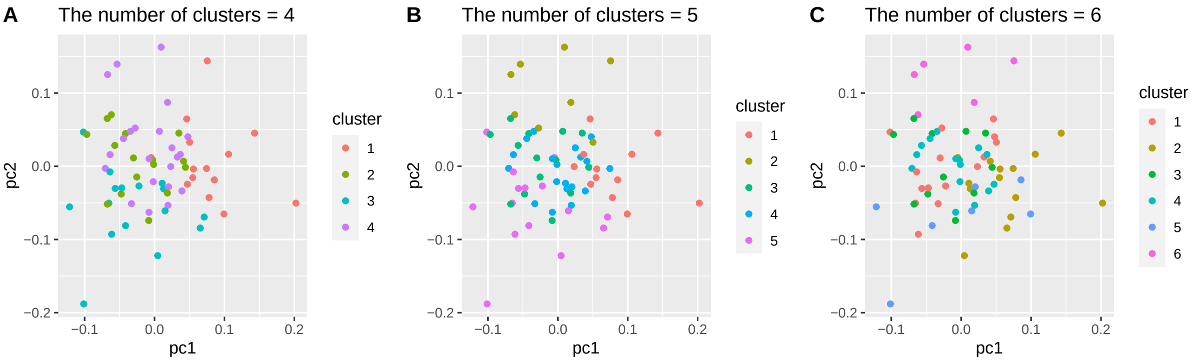
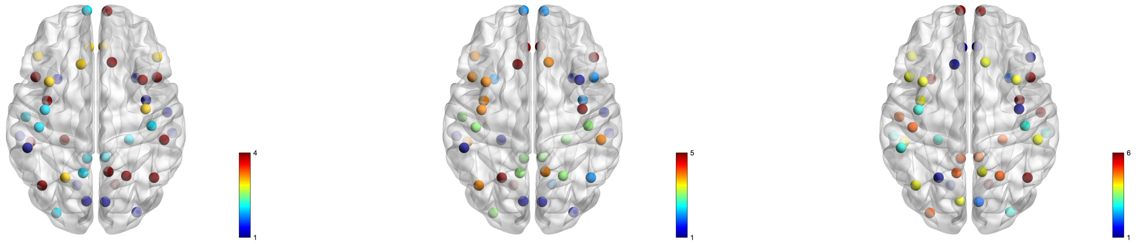


Figure 8: Clustering results according the number of clusters. The first axis is the largest principle vector and the second on is the second largest principle vector. A: when the number of clusters = 4, B: when the number of clusters = 5, and C: when the number of clusters = 6.



(a) The number of clusters = 4 (b) The number of clusters = 5 (c) The number of clusters = 6

Figure 9: Figure of brain nodes according to clusters when rank = 30.

3.3 Clustering results when rank = 49

The following is the elbow plot when rank = 49. The elbow plot does not give us the optimal group size because it decreases consistently with same rate according to the group size. So I stick to the group size 4,5, and 6 as before. When the group size is 7 or 8 are in Appendix.

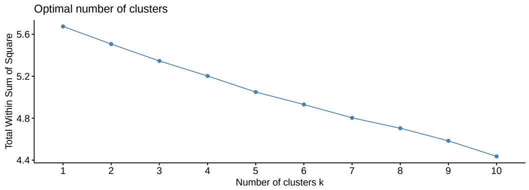


Figure 10: The elbow plot for determining the number of clusters in K-means.

I check the PCA results and brain image results for each case. Figure 11 shows brain nodes according to top two principle vectors and Figure 12 plots the brain nodes according to the clusters.

3.4 Clustering results when rank = 68

The following is the elbow plot when rank = 68. The elbow plot also does not give us the optimal group size because it decreases consistently with same rate according to the group size. So I stick

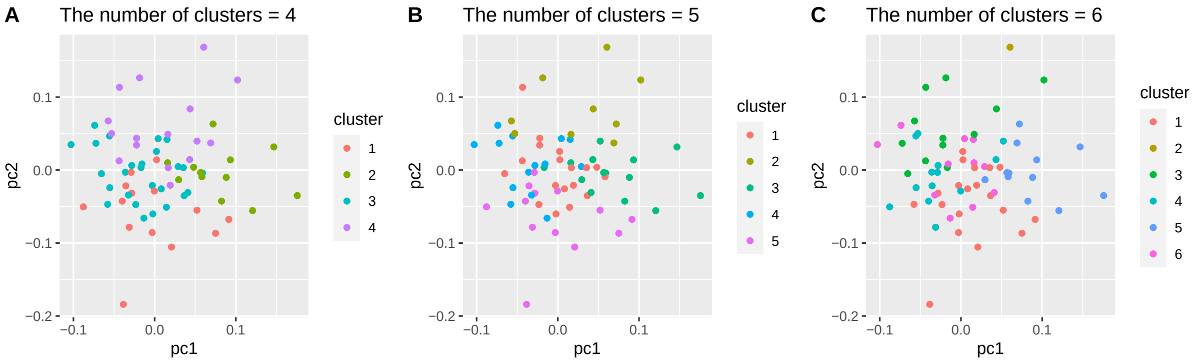
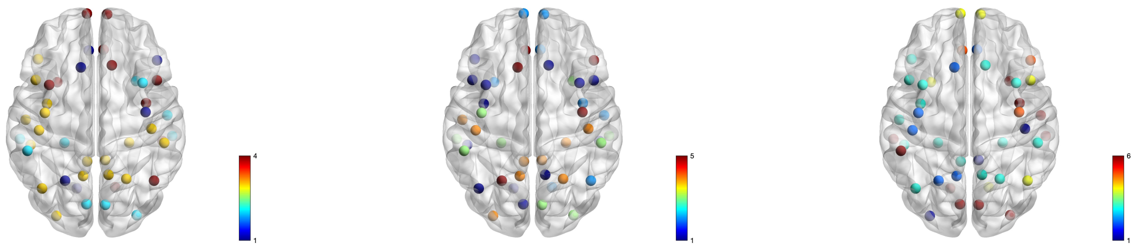


Figure 11: Clustering results according the number of clusters. The first axis is the largest principle vector and the second on is the second largest principle vector. A: when the number of clusters = 4, B: when the number of clusters = 5, and C: when the number of clusters = 6.



(a) The number of clusters = 4 (b) The number of clusters = 5 (c) The number of clusters = 6

Figure 12: Figure of brain nodes according to clusters when rank = 49.

to the group size 4,5, and 6 as before. When the group size is 7 or 8 are in Appendix.

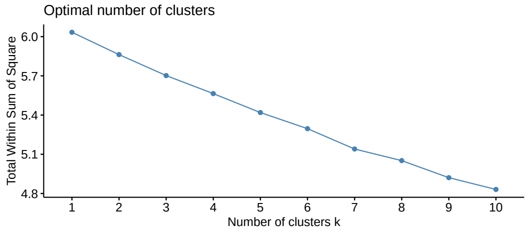


Figure 13: The elbow plot for determining the number of clusters in K-means.

I check the PCA results and brain image results for each case. Figure 14 shows brain nodes according to top two principle vectors and Figure 15 plots the brain nodes according to the clusters.

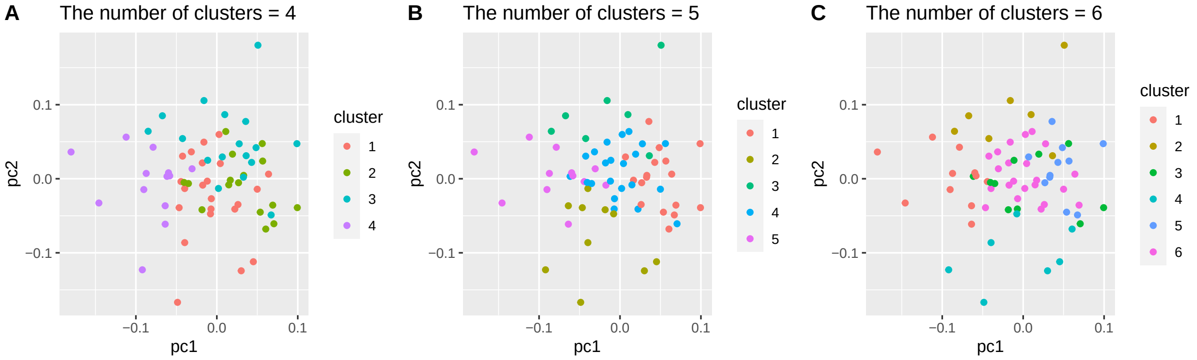
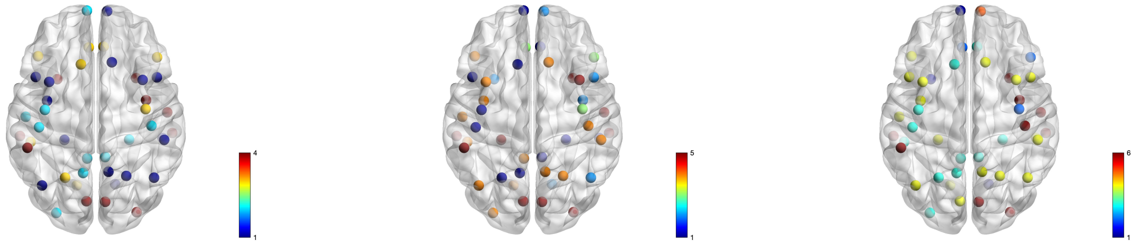


Figure 14: Clustering results according the number of clusters. The first axis is the largest principle vector and the second on is the second largest principle vector. A: when the number of clusters = 4, B: when the number of clusters = 5, and C: when the number of clusters = 6.



(a) The number of clusters = 4 (b) The number of clusters = 5 (c) The number of clusters = 6

Figure 15: Figure of brain nodes according to clusters when rank = 68 (full rank).

4 Probability estimation: IQ brain dataset

I perform weighted large margin classification with smooth parameter $H = 100$ and calculated probability $\mathbb{P}(Y = 1|\boldsymbol{X})$ when the rank = 8. However, the estimated probabilities were trivial when $\boldsymbol{X}$ is from the datasets. To be specific, when I calculate $\boldsymbol{X}_i$ for $i = 1, \dots, 114$, we have the following estimation.

$$\mathbb{P}(y_i|\boldsymbol{X}_i) = \begin{cases} 0.01 & \text{when observed } y_i = -1, \\ 0.99 & \text{when observed } y_i = 1. \end{cases}$$

Considering the precision of the estimation is 0.01 since $H = 100$, this result does not give us more information. One reason for we are having this estimation is that our model can fit the data perfectly when the rank = 8. Considering the number of parameters around 68×8 while available sample size is only 114, we can always find the parameter that makes 0-1 loss zero. This phenomenon can be explained by the following figure. Figure 16 shows that weighted classification results are the same regardless of the weight when the rank is greater than the certain rank. This same classification results in the trivial probability estimation.

Therefore, I think we need to reduce the rank size to avoid getting a trivial probability estimation. I changed the rank to 3 and am waiting for the results whether the probability estimation based on new rank gives us non-triivial information.

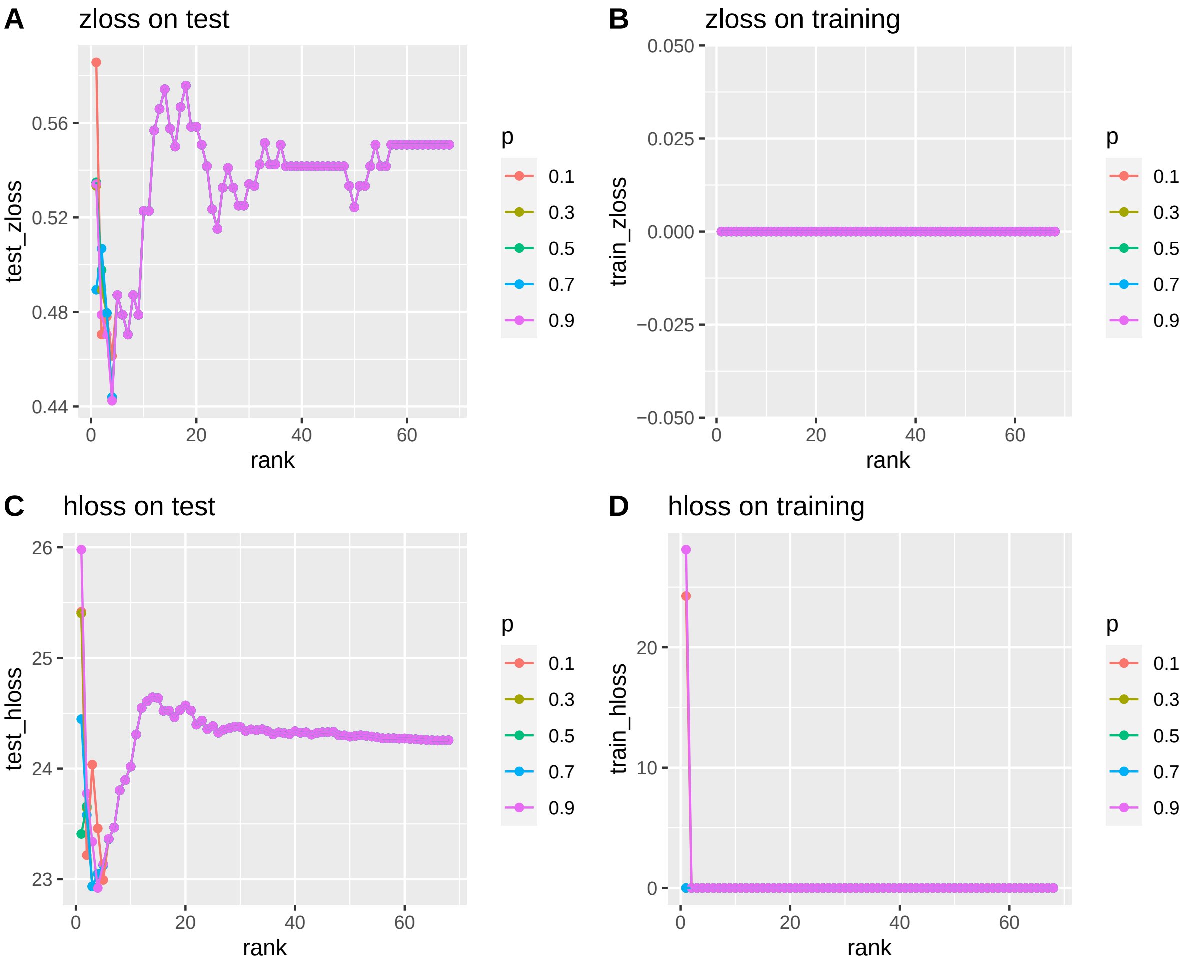


Figure 16: 5-folded cross validation results. A: 0-1 loss on test dataset, B: 0-1 loss on training dataset, C: hinge loss on test dataset, and D: hinge loss on training datset.

A Brain VSPLOT clustering results when group size is 7 or 8

B Probability estimation simulation

B.1 Sim1

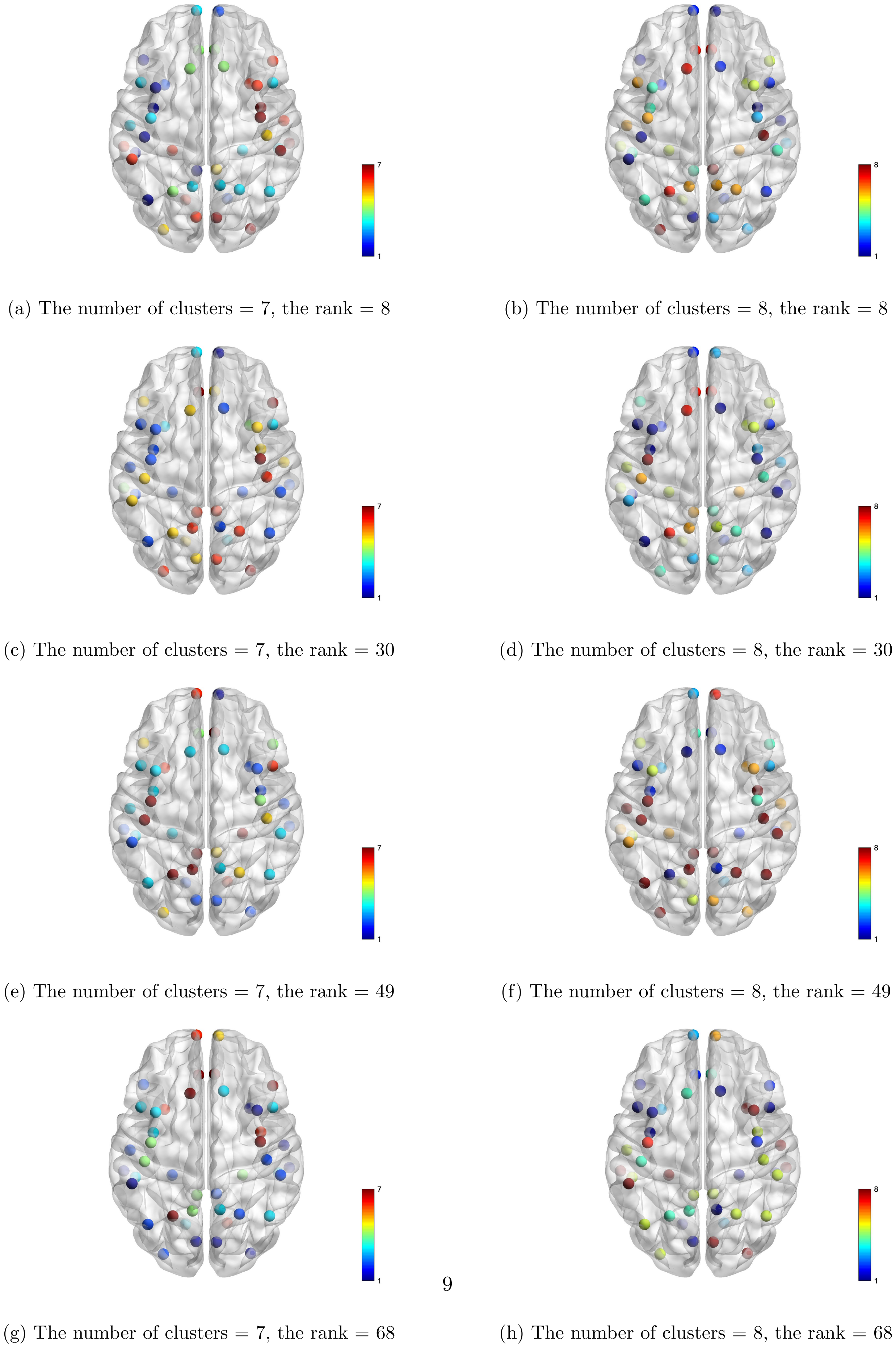
Model:

$$y = \operatorname{sign}\left(\langle \boldsymbol{B}, \boldsymbol{X} \rangle\right).$$

B.2 Sim2

Model: let $z_i = \langle \boldsymbol{B}, \boldsymbol{X}_i \rangle$,

$$y_i \sim \operatorname{Ber}\left(\frac{z_i \min_i z_i}{\max_i z_i - \min_i z_i}\right).$$



(a) The number of clusters = 7, the rank = 8

(b) The number of clusters = 8, the rank = 8

(c) The number of clusters = 7, the rank = 30

(d) The number of clusters = 8, the rank = 30

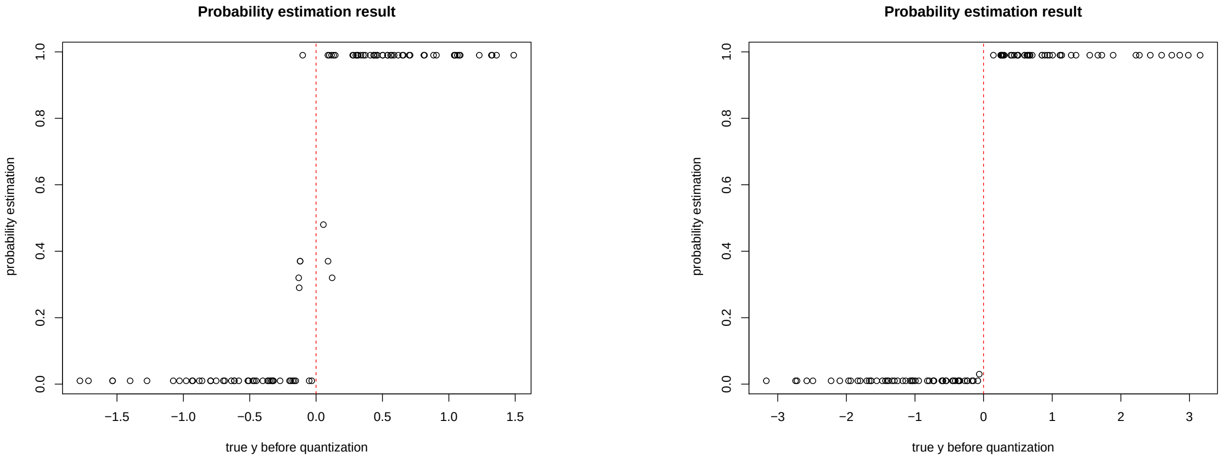
(e) The number of clusters = 7, the rank = 49

(f) The number of clusters = 8, the rank = 49

(g) The number of clusters = 7, the rank = 68

(h) The number of clusters = 8, the rank = 68

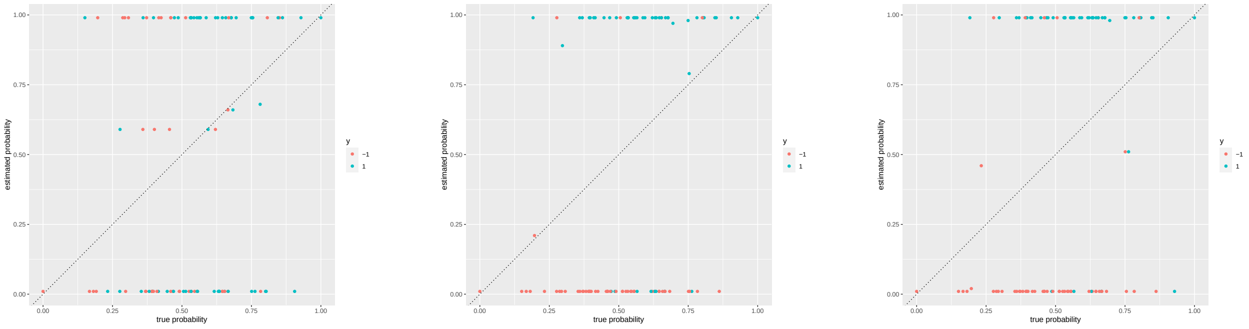
Figure 17: Figure of brain nodes according to clusters



(a) Linear kernel and binary feature

(b) linear kernel and continuous feature

Figure 18: Sim 1: No probability structure

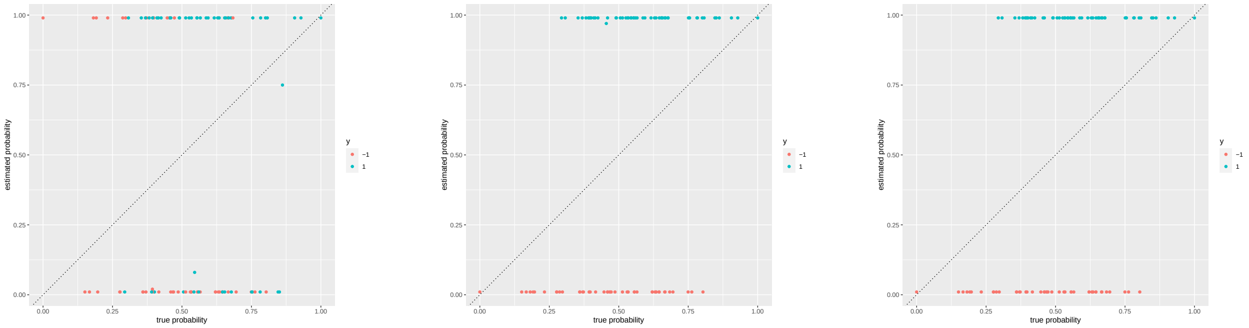


(a) Linear kernel

(b) Polynomial kernel

(c) Exponential kernel

Figure 19: Sim 2:When feature matrices are binary



(a) Linear kernel

(b) Polynomial kernel

(c) Exponential kernel

Figure 20: Sim 2:When feature matrices are continuous