Adovanced Data Analysis Homework 12

Information Science and Technology
Mathematical Informatics, M2
48226203 Kota Isayama *

July 31, 2023

Problem 1

We implement FDA. The source code is here[1].
The execution results are shown as follows:

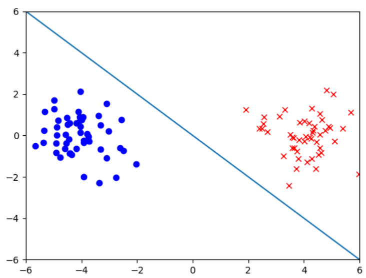

Figure 1: FDA with 2 classes, where there is no cluster structure within class.

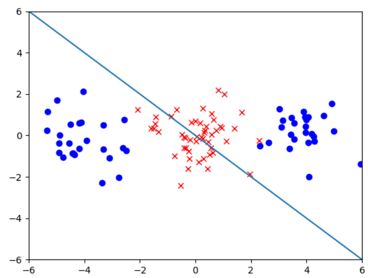

Figure 2: FDA with 2 class, where one of them consists of 2 separate sub clusters.

Both hyperplane can reduce dimensions of samples so that we can classify the data into correct classes.

*isakou2208@g.ecc.u-tokyo.ac.jp
[1]https://github.com/Kota-Isayama/Advanced-Data-Analysis/blob/main/12/problem1.py

Problem 2

We first derive the equation about $S^{(w)}$, then prove the one about $S^{(b)}$. We have

$$
\begin{aligned}
S^{(w)} &= \sum_{y=1}^{c} \sum_{i:y_i=y} (\boldsymbol{x}_i - \boldsymbol{\mu}_y)(\boldsymbol{x}_i - \boldsymbol{\mu}_y)^T \\
&= \sum_{y=1}^{c} \sum_{i:y_i=y} \left(\boldsymbol{x}_i \boldsymbol{x}_i^T - \boldsymbol{\mu}_y \boldsymbol{\mu}_y^T \right) \\
&= \sum_{y=1}^{c} \sum_{i:y_i=y} \boldsymbol{x}_i \boldsymbol{x}_i^T - \sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^T \\
&= \sum_{y=1}^{c} \sum_{i:y_i=y} \boldsymbol{x}_i \boldsymbol{x}_i^T - \sum_{y=1}^{c} \frac{1}{n_y} \left(\sum_{i:y_i=y} \boldsymbol{x}_i \right) \left(\sum_{i':y_{i'}=y} \boldsymbol{x}_{i'} \right)^T \\
&= \frac{1}{2} \sum_{y=1}^{c} \frac{1}{n_y} \left(\sum_{i:y_i=y} \sum_{i':y_{i'}=y} \left(\boldsymbol{x}_i \boldsymbol{x}_i^T + \boldsymbol{x}_{i'} \boldsymbol{x}_{i'}^T - 2 \boldsymbol{x}_i \boldsymbol{x}_{i'}^T \right) \right) \\
&= \frac{1}{2} \sum_{y=1}^{c} \sum_{i:y_i=y} \sum_{i':y_{i'}=y} \frac{1}{n_y} (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^2 \\
&= \frac{1}{2} \sum_{i,i'=1}^{n} Q^{(w)}_{i,i'} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^T ,
\end{aligned}
$$

where 4th equality is due to the definition of $\boldsymbol{\mu}_y$, 5th equality holds since $\sum_{i:y_i=y} \boldsymbol{x}_i \boldsymbol{x}_i^T = \frac{1}{n_y} \sum_{i:y_i=y} \sum_{i':y_{i'}=y} \boldsymbol{x}_i \boldsymbol{x}_i^T$ and $\sum_{i:y_i=y} \boldsymbol{x}_i \boldsymbol{x}_i^T = \frac{1}{2} \left(\sum_{i:y_i=y} \boldsymbol{x}_i \boldsymbol{x}_i^T + \sum_{i':y_{i'}=y} \boldsymbol{x}_{i'} \boldsymbol{x}_{i'}^T \right)$ and 7th equality is derived from the definition of $Q^{(w)}_{i,i'}$.

Then, we prove the equation about $S^{(b)}$. From $C = S^{(w)} + S^{(b)}$, where $C = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$, we have

$$
\begin{aligned}
S^{(b)} &= C - S^{(w)} \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T - \frac{1}{2} \sum_{i,i'=1}^{n} Q^{(w)}_{i,i'} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^T \\
&= \frac{1}{2} \sum_{i,i'=1}^{n} \frac{1}{n} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^T - \frac{1}{2} \sum_{i,i'=1}^{n} Q^{(w)}_{i,i'} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^T \\
&= \frac{1}{2} \sum_{i,i'=1}^{n} Q^{(b)}_{i,i'} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^T ,
\end{aligned}
$$

where third equation holds since

$$
\begin{aligned}
\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T &= \frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} \left(\boldsymbol{x}_i \boldsymbol{x}_i^T + \boldsymbol{x}_{i'} \boldsymbol{x}_{i'}^T \right) \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^T \qquad (\because \sum_{i=1}^{n} \boldsymbol{x}_i = \boldsymbol{0}).
\end{aligned}
$$

.

Problem 3

We show $\mathrm{rank}(S^{(b)}) \leq c - 1$, where c is the number of classes and $\sum_{i=1}^{n} \boldsymbol{x}_i = \boldsymbol{0}$. Since $\sum_{i=1}^{n} \boldsymbol{x}_i = \boldsymbol{0}$, we have

$$
\begin{aligned}
\sum_{y=1}^{c} n_y \boldsymbol{\mu}_y &= \boldsymbol{0} \\
n_c \boldsymbol{\mu}_c &= - \sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y.
\end{aligned}
$$

Then, from the definition of $S^{(b)}$,

$$
\begin{aligned}
S^{(b)} &= \sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^T \\
&= \sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^T + n_c \boldsymbol{\mu}_c \boldsymbol{\mu}_c^T \\
&= \sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^T - \frac{1}{n_c} \left(- \sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y \right) \left(- \sum_{y'=1}^{c-1} n_{y'} \boldsymbol{\mu}_{y'} \right)^T \\
&= \sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y \left(\boldsymbol{\mu}_y + \frac{1}{n_c} \sum_{y'=1}^{c-1} n_{y'} \boldsymbol{\mu}_{y'} \right)^T \\
&= \sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y \boldsymbol{z}_y^T,
\end{aligned}
$$

where $\boldsymbol{z}_y = \boldsymbol{\mu}_y + \frac{1}{n_c} \sum_{y'=1}^{c-1} n_{y'} \boldsymbol{\mu}_{y'}$ is a vector. Generally, for any two vectors $\boldsymbol{a}, \boldsymbol{b}$ of same dimension, $\mathrm{rank}(\boldsymbol{a}\boldsymbol{b}^T) \leq 1$. This is because $\boldsymbol{a}\boldsymbol{b}^T$ can be represented as $(b_1 \boldsymbol{a}, b_2 \boldsymbol{a}, \ldots, b_d \boldsymbol{a})$, where $\boldsymbol{b} = (b_1, b_2, \ldots, b_d)^T$ and all $b_i \boldsymbol{a}$ have same direction. From the above, $\mathrm{rank}(n_y \boldsymbol{\mu}_y \boldsymbol{z}_y^T) \leq$

1 for all y. Finally, we have

$$\begin{aligned}\operatorname{rank}\left(S^{(b)}\right) &= \operatorname{rank}\left(\sum_{y=1}^{c-1} n_y \boldsymbol{\mu}_y \boldsymbol{z}_y^T\right) \\ &\leq \sum_{y=1}^{c-1} \operatorname{rank}\left(n_y \boldsymbol{\mu}_y \boldsymbol{z}_y^T\right) \\ &\leq c-1,\end{aligned}$$

where first inequality holds due to the subadditivity of rank of matrices.

Problem 4

We implement least squares probabilistic classification based on a Gaussian kernel model:

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) = \sum_{i:y_i=y} \theta_i^{(y)} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|}{2h^2}\right).$$

The source code is here[2].

In figure3, we show result when the number of classes is 3, the parameter $h = 2.0$ and the training sample size is 90.

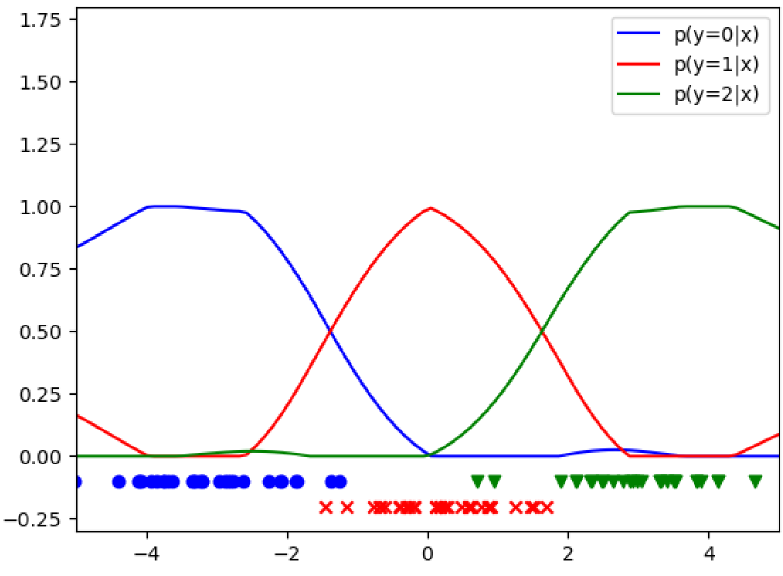

Figure 3: The least squares probabilistic classification based on a Gaussian kernel model, where $h = 2.0$.

[2]https://github.com/Kota-Isayama/Advanced-Data-Analysis/blob/main/12/problem4.py