Homework 2 STOR 767: Theory Part

Due Sep 27, 2020

Student Name: Leo Li (PID: 730031954)

Instructions

- Edit this LaTeX file with your solutions and generate a PDF file from it. Upload **both the tex and the pdf** file to Sakai.
- Use proper fonts for a clear presentation:
  x for an observed value; X for a random variable; $\boldsymbol{X}$ for a vector; $\mathbf{X}$ for a matrix.
- You are allowed to work with other students but homework should be in your own words. Identical solutions will receive a **0** in grade and will be investigated.

1. Consider i.i.d. observations $(\boldsymbol{X}_i, Y_i), i = 1, \ldots, n$ from the QDA model: $Y \sim Bernoulli(\eta), \boldsymbol{X}|Y = +1 \sim \mathcal{N}(\boldsymbol{\mu}_+, \Sigma_+), \boldsymbol{X}|Y = -1 \sim \mathcal{N}(\boldsymbol{\mu}_-, \Sigma_-)$. Find the MLE of the parameters $(\eta, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \Sigma_+, \Sigma_-)$.

Solution.

In this question, we would like to derive the MLE for the parameters $(\eta, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \Sigma_+, \Sigma_-)$ in the QDA model. First, we start by writing out the likelihood function as the following:

$$L_n = \prod_{i:y_i=+1} \eta \frac{1}{(2\pi)^{d/2}|\Sigma_+|^{1/2}} \exp[-\frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma_+^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2}]$$
$$\prod_{i:y_i=-1} (1-\eta) \frac{1}{(2\pi)^{d/2}|\Sigma_-|^{1/2}} \exp[-\frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T \Sigma_-^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2}].$$

Let $n_+ = \sum_{i=1}^n I(y_i = +1)$ and $n_- = \sum_{i=1}^n I(y_i = -1)$. Then, the log likelihood function can be derived as:

$$l_n = n_+ \log \eta - \frac{dn_+}{2}\log(2\pi) - \frac{n_+}{2}\log|\Sigma_+| - \sum_{i:y_i=+1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma_+^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2}$$
$$n_- \log(1-\eta) - \frac{dn_-}{2}\log(2\pi) - \frac{n_-}{2}\log|\Sigma_-| - \sum_{i:y_i=-1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T \Sigma_-^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2}.$$

Then, we start to derive the MLE for the parameters. For η, the score function is that,

$$\frac{\partial}{\partial \eta} l_n = \frac{n_+}{\eta} - \frac{n_-}{1-\eta}.$$

The MLE of η can be obtained by setting the score function to zero, and we will get that,

$$\hat{\eta} = \frac{n_+}{n_+ + n_-}.$$

For $\boldsymbol{\mu}_+$, the score function is that,

$$\frac{\partial}{\partial \boldsymbol{\mu}_+} l_n = \sum_{i:y_i=+1} \Sigma_+^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+).$$

The MLE of $\boldsymbol{\mu}_+$ can be obtained by setting the score function to zero, and we will get that,

$$\hat{\boldsymbol{\mu}}_+ = \frac{1}{n_+} \sum_{i:y_i=+1} \boldsymbol{X}_i.$$

Now, we would like to derive the MLE for Σ_+ by re-expressing the log likelihood function as the following way:

$$l_n = C - \frac{n_+}{2}\log|\Sigma_+| - \sum_{i:y_i=+1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma_+^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2}$$
$$= C + \frac{n_+}{2}\log|\Sigma_+^{-1}| - \frac{1}{2}\sum_{i:y_i=+1} tr[(\boldsymbol{X}_i - \boldsymbol{\mu}_+)(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma_+^{-1}].$$

Then the score equation can be derived as,

$$\frac{\partial}{\partial \Sigma_+^{-1}} l_n = \frac{n_+}{2}\Sigma_+ - \frac{1}{2}\sum_{i:y_i=+1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+)(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T.$$

The MLE of Σ_+ can be obtained by setting the score function to zero, and by invariance property of MLE, we will get that,

$$\hat{\Sigma}_+ = \frac{1}{n_+}\sum_{i:y_i=+1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)^T.$$

By using the same reasoning, we can also derive the MLE of $\boldsymbol{\mu}_-$ and Σ_- as the following:

$$\hat{\boldsymbol{\mu}}_- = \frac{1}{n_-}\sum_{i:y_i=-1}\boldsymbol{X}_i,$$

$$\hat{\Sigma}_- = \frac{1}{n_-}\sum_{i:y_i=-1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)^T.$$

To sum up, we find the MLE of the parameters for the QDA model as the following:

$$\begin{bmatrix} \hat{\eta} \\ \hat{\boldsymbol{\mu}}_+ \\ \hat{\boldsymbol{\mu}}_- \\ \hat{\Sigma}_+ \\ \hat{\Sigma}_- \end{bmatrix} = \begin{bmatrix} \frac{n_+}{n_+ + n_-} \\ \frac{1}{n_+}\sum_{i:y_i=+1}\boldsymbol{X}_i \\ \frac{1}{n_-}\sum_{i:y_i=-1}\boldsymbol{X}_i \\ \frac{1}{n_+}\sum_{i:y_i=+1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)^T \\ \frac{1}{n_-}\sum_{i:y_i=-1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)^T \end{bmatrix}.$$

2. Consider i.i.d. observations $(\boldsymbol{X}_i, Y_i), i = 1, \ldots, n$ from the LDA model: $Y \sim Bernoulli(\eta), \boldsymbol{X}|Y = +1 \sim \mathcal{N}(\boldsymbol{\mu}_+, \Sigma), \boldsymbol{X}|Y = -1 \sim \mathcal{N}(\boldsymbol{\mu}_-, \Sigma)$. Find the MLE of the parameters $(\eta, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \Sigma)$.

SOLUTION.

In this question, we would like to derive the MLE for the parameters $(\eta, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \Sigma)$ in the LDA model. First, we start by writing out the likelihood function as the following:

$$L_n = \prod_{i:y_i=+1} \eta \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2}]$$
$$\prod_{i:y_i=-1} (1-\eta) \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2}].$$

Let $n_+ = \sum_{i=1}^n I(y_i = +1)$ and $n_- = \sum_{i=1}^n I(y_i = -1)$. Then, the log likelihood function can be derived as:

$$l_n = n_+ \log \eta - \frac{dn_+}{2} \log(2\pi) - \frac{n_+}{2} \log|\Sigma| - \sum_{i:y_i=+1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2}$$
$$n_- \log(1-\eta) - \frac{dn_-}{2} \log(2\pi) - \frac{n_-}{2} \log|\Sigma| - \sum_{i:y_i=-1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2}.$$

Then, we start to derive the MLE for the parameters. For η, the score function is that,

$$\frac{\partial}{\partial \eta} l_n = \frac{n_+}{\eta} - \frac{n_-}{1-\eta}.$$

The MLE of η can be obtained by setting the score function to zero, and we will get that,

$$\hat{\eta} = \frac{n_+}{n_+ + n_-}.$$

For $\boldsymbol{\mu}_+$, the score function is that,

$$\frac{\partial}{\partial \boldsymbol{\mu}_+} l_n = \sum_{i:y_i=+1} \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_+).$$

The MLE of $\boldsymbol{\mu}_+$ can be obtained by setting the score function to zero, and we will get that,

$$\hat{\boldsymbol{\mu}}_+ = \frac{1}{n_+} \sum_{i:y_i=+1} \boldsymbol{X}_i.$$

By using similar reasoning, we can obtain the MLE of $\boldsymbol{\mu}_-$ as,

$$\hat{\boldsymbol{\mu}}_- = \frac{1}{n_-} \sum_{i:y_i=-1} \boldsymbol{X}_i.$$

Now, we would like to derive the MLE for Σ by re-expressing the log likelihood function as the following way:

$$
\begin{aligned}
l_n &= C - \frac{n}{2}\log|\Sigma| - \sum_{i:y_i=+1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T\Sigma^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2} \\
&\quad - \sum_{i:y_i=-1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T\Sigma^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2} \\
&= C + \frac{n}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{i:y_i=+1} tr[(\boldsymbol{X}_i - \boldsymbol{\mu}_+)(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T\Sigma^{-1}] \\
&\quad - \frac{1}{2}\sum_{i:y_i=-1} tr[(\boldsymbol{X}_i - \boldsymbol{\mu}_-)(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T\Sigma^{-1}].
\end{aligned}
$$

Then the score equation can be derived as,

$$
\frac{\partial}{\partial\Sigma^{-1}} l_n = \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i:y_i=+1}(\boldsymbol{X}_i - \boldsymbol{\mu}_+)(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T - \frac{1}{2}\sum_{i:y_i=-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_-)(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T.
$$

The MLE of Σ can be obtained by setting the score function to zero, and by invariance property of MLE, we will get that,

$$
\hat{\Sigma} = \frac{1}{n}\{\sum_{i:y_i=+1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)^T + \sum_{i:y_i=-1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)^T\}.
$$

To sum up, we find the MLE of the parameters for the LDA model as the following:

$$
\begin{bmatrix} \hat{\eta} \\ \hat{\boldsymbol{\mu}}_+ \\ \hat{\boldsymbol{\mu}}_- \\ \hat{\Sigma} \end{bmatrix} = \begin{bmatrix} \frac{n_+}{n_+ + n_-} \\ \frac{1}{n_+}\sum_{i:y_i=+1}\boldsymbol{X}_i \\ \frac{1}{n_-}\sum_{i:y_i=-1}\boldsymbol{X}_i \\ \frac{1}{n}\{\sum_{i:y_i=+1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_+)^T + \sum_{i:y_i=-1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_-)^T\} \end{bmatrix}.
$$

3. Consider i.i.d. observations $(\boldsymbol{X}_i, Y_i), i = 1, \ldots, n$ from the DLDA model: $Y \sim Bernoulli(\eta), \boldsymbol{X}|Y = +1 \sim \mathcal{N}(\boldsymbol{\mu}_+, \Sigma), \boldsymbol{X}|Y = -1 \sim \mathcal{N}(\boldsymbol{\mu}_-, \Sigma)$, where Σ is a diagonal matrix with diagonal entries $\sigma_j^2, j = 1, \ldots, d$. Find the MLE of the parameters $(\eta, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \sigma_1^2, \ldots, \sigma_d^2)$.

Solution.

In this question, we would like to derive the MLE for the parameters $(\eta, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \sigma_1^2, \ldots, \sigma_d^2)$ in the DLDA model. First, we start by writing out the likelihood function as the following:

$$
\begin{aligned}
L_n = &\prod_{i:y_i=+1} \eta \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2}] \\
&\prod_{i:y_i=-1} (1-\eta) \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_-)^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2}].
\end{aligned}
$$

Let $n_+ = \sum_{i=1}^n I(y_i = +1)$ and $n_- = \sum_{i=1}^n I(y_i = -1)$. Then, the log likelihood function can be derived as:

$$
\begin{aligned}
l_n = n_+ \log \eta - \frac{dn_+}{2} \log(2\pi) - \frac{n_+}{2} \log|\Sigma| - \sum_{i:y_i=+1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu}_+)^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_+)}{2} \\
n_- \log(1-\eta) - \frac{dn_-}{2} \log(2\pi) - \frac{n_-}{2} \log|\Sigma| - \sum_{i:y_i=-1} \frac{(\boldsymbol{X}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_-)}{2}.
\end{aligned}
$$

Then, we start to derive the MLE for the parameters. For η, the score function is that,

$$\frac{\partial}{\partial \eta} l_n = \frac{n_+}{\eta} - \frac{n_-}{1-\eta}.$$

The MLE of η can be obtained by setting the score function to zero, and we will get that,

$$\hat{\eta} = \frac{n_+}{n_+ + n_-}.$$

For $\boldsymbol{\mu}_+$, the score function is that,

$$\frac{\partial}{\partial \boldsymbol{\mu}_+} l_n = \sum_{i:y_i=+1} \Sigma^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}_+).$$

The MLE of $\boldsymbol{\mu}_+$ can be obtained by setting the score function to zero, and we will get that,

$$\hat{\boldsymbol{\mu}}_+ = \frac{1}{n_+} \sum_{i:y_i=+1} \boldsymbol{X}_i.$$

By using similar reasoning, we can obtain the MLE of $\boldsymbol{\mu}_-$ as,

$$\hat{\boldsymbol{\mu}}_- = \frac{1}{n_-} \sum_{i:y_i=-1} \boldsymbol{X}_i.$$

Now, we would like to derive the MLE of $(\sigma_1^2, \dots, \sigma_d^2)$. Since we assume that Σ is a diagonal matrix, then we can rewrite the likelihood function as the following:

$$
\begin{aligned}
L_n = & \prod_{i:y_i=+1} \eta \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \prod_{j=1}^{d} \exp[-\frac{(X_{ij}-\mu_{+j})^2}{2\sigma_j^2}] \\
& \qquad \prod_{i:y_i=-1} (1-\eta) \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \prod_{j=1}^{d} \exp[-\frac{(X_{ij}-\mu_{-j})^2}{2\sigma_j^2}] \\
= & \prod_{i:y_i=+1} \eta \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\sum_{j=1}^{d} \frac{(X_{ij}-\mu_{+j})^2}{2\sigma_j^2}] \\
& \qquad \prod_{i:y_i=-1} (1-\eta) \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\sum_{j=1}^{d} \frac{(X_{ij}-\mu_{-j})^2}{2\sigma_j^2}],
\end{aligned}
$$

where X_{ij} represents the j^{th} element of $\boldsymbol{X}_i$, μ_{+j} represents the j^{th} element of $\boldsymbol{\mu}_+$, and μ_{-j} represents the j^{th} element of $\boldsymbol{\mu}_-$. Then the log likelihood function can be written as,

$$
\begin{aligned}
l_n = C + \frac{n}{2}\log(\prod_{j=1}^{d}\sigma_j^{-2}) - \frac{1}{2}\sum_{i:y_i=+1}[\sum_{j=1}^{d}\frac{(X_{ij}-\mu_{+j})^2}{\sigma_j^2}] \\
- \frac{1}{2}\sum_{i:y_i=-1}[\sum_{j=1}^{d}\frac{(X_{ij}-\mu_{-j})^2}{\sigma_j^2}].
\end{aligned}
$$

Then the score function can be derived as,

$$
\frac{\partial}{\partial \sigma_j^2} l_n = \frac{n}{2\sigma_j^{-2}} - \frac{1}{2}\{\sum_{i:y_i=+1}(X_{ij}-\mu_{+j})^2 + \sum_{i:y_i=-1}(X_{ij}-\mu_{-j})^2\}.
$$

The MLE of σ_j^2, where $j = 1, \dots, d$, can be obtained by setting the score function to zero, and by invariance property of MLE, we will get that,

$$
\hat{\sigma}_j^2 = \frac{1}{n}\{\sum_{i:y_i=+1}(X_{ij}-\hat{\mu}_{+j})^2 + \sum_{i:y_i=-1}(X_{ij}-\hat{\mu}_{-j})^2\}.
$$

To sum up, we find the MLE of the parameters for the DLDA model as the following:

$$
\begin{bmatrix} \hat{\eta} \\ \hat{\boldsymbol{\mu}}_+ \\ \hat{\boldsymbol{\mu}}_- \\ \hat{\sigma}_j^2 \end{bmatrix} = \begin{bmatrix} \frac{n_+}{n_+ + n_-} \\ \frac{1}{n_+}\sum_{i:y_i=+1}\boldsymbol{X}_i \\ \frac{1}{n_-}\sum_{i:y_i=-1}\boldsymbol{X}_i \\ \frac{1}{n}\{\sum_{i:y_i=+1}(X_{ij}-\hat{\mu}_{+j})^2 + \sum_{i:y_i=-1}(X_{ij}-\hat{\mu}_{-j})^2\} \end{bmatrix},
$$

where $j = 1, \dots, d$.