

Gaussian Setting Detection Proofs

June 13, 2011

Vector Case with Estimated Mean

We consider the following binary classification problem:

$$y = \begin{cases} \mu_0 + z & y \in H_0 \\ \mu_1 + z & y \in H_1 \end{cases} \quad (1)$$

where $z \in \mathbf{R}^n, z \sim \mathcal{N}(0, I)$. We assume that $\mu_0, \mu_1 \in \mathbf{R}^n$ are unknown but we have estimates $\hat{\mu}_0, \hat{\mu}_1$ which were estimated from the data. We assume that $f(\mu_0|\hat{\mu}_0) \sim \mathcal{N}(0, \sigma_0^2 I)$ and $f(\mu_1|\hat{\mu}_1) \sim \mathcal{N}(1, \sigma_1^2 I)$

Let $\tilde{\mu}_0 = \mu_0|\hat{\mu}_0$ and $\tilde{\mu}_1 = \mu_1|\hat{\mu}_1$ so our new classification problem becomes

$$y = \begin{cases} \tilde{\mu}_0 + z & y \in H_0 \\ \tilde{\mu}_1 + z & y \in H_1 \end{cases} \quad (2)$$

where $z \sim \mathcal{N}(0, I), \tilde{\mu}_0 \sim \mathcal{N}(0, \sigma_0^2 I), \tilde{\mu}_1 \sim \mathcal{N}(0, \sigma_1^2 I)$.

Under $H_0, y \sim \mathcal{N}(0, (1 + \sigma_0^2)I)$ and under $H_1, y \sim \mathcal{N}(0, (1 + \sigma_1^2)I)$. Our LRT is

$$\begin{aligned} \Lambda(y) = \frac{f_1(y)}{f_0(y)} &= \frac{(2\pi(1 + \sigma_1^2))^{-n/2} \exp\{\frac{-1}{2(1+\sigma_1^2)} y^T y\}}{(2\pi(1 + \sigma_0^2))^{-n/2} \exp\{\frac{-1}{2(1+\sigma_0^2)} y^T y\}} \\ &= \left(\frac{1 + \sigma_0^2}{1 + \sigma_1^2} \right)^{n/2} \exp \left\{ \frac{-y^T y}{2} \left(\frac{1}{1 + \sigma_1^2} - \frac{1}{1 + \sigma_0^2} \right) \right\} \\ &= \left(\frac{1 + \sigma_0^2}{1 + \sigma_1^2} \right)^{n/2} \exp \left\{ \frac{-y^T y}{2} \left(\frac{\sigma_0^2 - \sigma_1^2}{(1 + \sigma_1^2)(1 + \sigma_0^2)} \right) \right\} \end{aligned} \quad (3)$$

Where our decision is

$$\begin{aligned} &\text{Declare } H_0 \text{ if } \Lambda(y) < \eta \\ &\text{Declare } H_1 \text{ if } \Lambda(y) > \eta \end{aligned} \quad (4)$$

Defining

$$\hat{\Lambda}(y) = \frac{y^T y}{2} \left(\frac{\sigma_1^2 - \sigma_0^2}{(1 + \sigma_1^2)(1 + \sigma_0^2)} \right) \quad (5)$$

our decision becomes

$$\begin{aligned} &\text{Declare } H_0 \text{ if } \hat{\Lambda}(y) < \ln \left(\eta \left(\frac{1 + \sigma_1^2}{1 + \sigma_0^2} \right)^{n/2} \right) \\ &\text{Declare } H_1 \text{ if } \hat{\Lambda}(y) > \ln \left(\eta \left(\frac{1 + \sigma_1^2}{1 + \sigma_0^2} \right)^{n/2} \right) \end{aligned} \quad (6)$$

We assume that $\sigma_1^2 > \sigma_0^2$. If this is not the case, we may simply reverse the roles of H_0, H_1 . Defining

$$\gamma = \frac{2(1 + \sigma_1^2)(1 + \sigma_0^2)}{\sigma_1^2 - \sigma_0^2} \left(\ln(\eta) + (n/2) \ln \left(\frac{1 + \sigma_1^2}{1 + \sigma_0^2} \right) \right) \quad (7)$$

Our decision becomes

$$\begin{aligned} \text{Declare } H_0 & \text{ if } y^T y < \gamma \\ \text{Declare } H_1 & \text{ if } y^T y > \gamma \end{aligned} \quad (8)$$

In the case where each class is equally likely ($\eta = 1$), this reduces to

$$\hat{\gamma} = \frac{n(1 + \sigma_1^2)(1 + \sigma_0^2)}{\sigma_1^2 - \sigma_0^2} \ln \left(\frac{1 + \sigma_1^2}{1 + \sigma_0^2} \right) \quad (9)$$

$$\begin{aligned} \text{Declare } H_0 & \text{ if } y^T y < \hat{\gamma} \\ \text{Declare } H_1 & \text{ if } y^T y > \hat{\gamma} \end{aligned} \quad (10)$$

Define our test statistic as $\tau = y^T y$. We seek the distribution of τ under H_0 . Now $\tau = \sum_{i=1}^n y_i^2$. Under H_0 , $y_i \sim \mathcal{N}(0, \sigma_0^2)$. Therefore under H_0 ,

$$\frac{\tau}{\sigma_0^2} = \sum_{i=1}^n \left(\frac{y_i}{\sigma_0} \right)^2 \quad (11)$$

which is a sum of standard normal Gaussians. Therefore, $\frac{\tau}{\sigma_0^2} \sim \chi_n^2$. Similarly, under H_1 , $\frac{\tau}{\sigma_1^2} \sim \chi_n^2$. We can now set γ to achieve a false alarm right α .

$$P_F = P(\tau > \gamma | \text{under } H_0) = P\left(\frac{\tau}{\sigma_0^2} > \frac{\gamma}{\sigma_0^2} | \text{under } H_0\right) = Q_{\chi_n^2}\left(\frac{\gamma}{\sigma_0^2}\right) \quad (12)$$

We can then solve for γ

$$\gamma = \sigma_0^2 Q_{\chi_n^2}^{-1}(P_F) \quad (13)$$

which we can then use to develop a relationship between P_F and P_D

$$P_D = P(\tau > \gamma | \text{under } H_1) = P\left(\frac{\tau}{\sigma_1^2} > \frac{\gamma}{\sigma_1^2} | \text{under } H_1\right) = Q_{\chi_n^2}\left(\frac{\gamma}{\sigma_1^2}\right) \quad (14)$$

Substituting our expression for γ in (13) we have

$$P_D = Q_{\chi_n^2}\left(\frac{\sigma_0^2}{\sigma_1^2} Q_{\chi_n^2}^{-1}(P_F)\right) \quad (15)$$

Processed Observation with Estimated Mean

We now consider the case where we are not given y explicitly but instead a processed version of it. That is

$$t = \begin{cases} U_0^T y & t \in H_0 \\ U_1^T y & t \in H_1 \end{cases} \quad (16)$$

where $U_1, U_2 \in \mathbf{R}^{n \times p}$ are known have orthonormal columns and $t \in \mathbf{R}^p$. We therefore have the detection problem

$$t = \begin{cases} U_0^T \mu_0 + U_0^T z & t \in H_0 \\ U_1^T \mu_1 + U_1^T z & t \in H_1 \end{cases} \quad (17)$$

where $z \in \mathbf{R}^n, z \sim \mathcal{N}(0, I)$. We assume that $\mu_0, \mu_1 \in \mathbf{R}^n$ are unknown but we have estimates $\hat{\mu}_0, \hat{\mu}_1$ which were estimated from the data. We assume that $f(\mu_0|\hat{\mu}_0) \sim \mathcal{N}(0, \sigma_0^2 I)$ and $f(\mu_1|\hat{\mu}_1) \sim \mathcal{N}(0, \sigma_1^2 I)$. As U_0, U_1 have orthonormal columns, $U_0 z, U_1 z$ are both $\mathcal{N}(0, I_{p \times p})$. Let $\tilde{z} \sim \mathcal{N}(0, I_{p \times p})$. Let $\tilde{\mu}_0 = \mu_0|\hat{\mu}_0$ and $\tilde{\mu}_1 = \mu_1|\hat{\mu}_1$ so our new classification problem becomes

$$t = \begin{cases} U_0^T \tilde{\mu}_0 + \tilde{z} & t \in H_0 \\ U_1^T \tilde{\mu}_1 + \tilde{z} & t \in H_1 \end{cases} \quad (18)$$

Under $H_0, t \sim \mathcal{N}(0, \sigma_0^2 I_{p \times p})$ and under $H_1, t \sim \mathcal{N}(0, \sigma_1^2 I_{p \times p})$

Similar to the previous section, this yields the detector

$$\begin{aligned} \text{Declare } H_0 \text{ if } t^T t &< \frac{2(1 + \sigma_1^2)(1 + \sigma_0^2)}{\sigma_1^2 - \sigma_0^2} \left(\ln(\eta) + (p/2) \ln \left(\frac{1 + \sigma_1^2}{1 + \sigma_0^2} \right) \right) \\ \text{Declare } H_1 \text{ if } t^T t &> \frac{2(1 + \sigma_1^2)(1 + \sigma_0^2)}{\sigma_1^2 - \sigma_0^2} \left(\ln(\eta) + (p/2) \ln \left(\frac{1 + \sigma_1^2}{1 + \sigma_0^2} \right) \right) \end{aligned} \quad (19)$$

Using a similar analysis to the previous section, we can establish a relationship between P_D and P_F using the distribution of our test statistic $t^T t$ and arrive at

$$P_D = Q_{\chi_p^2} \left(\frac{\sigma_0^2}{\sigma_1^2} Q_{\chi_p^2}^{-1}(P_F) \right) \quad (20)$$

with the only difference being the degrees of freedom of our chi-square distribution.

Numerical Results

We plot the theoretical ROC curves derived in each of the previous two sections. The only difference between the two being the dimension of the test statistic which results in different degrees of freedom in the resulting chi-square distribution. We can see that when we reduce the dimension of our problem, we push the ROC curve to the bottom right which is evidence of a worse detector. The larger the ratio of n/p results in a larger decrease in detection power. The larger the ratio of σ_1^2/σ_0^2 results in overall better detectors (for both detectors). This intuitively makes sense as our observed points will be clustered more sparsely and easier to differentiate.

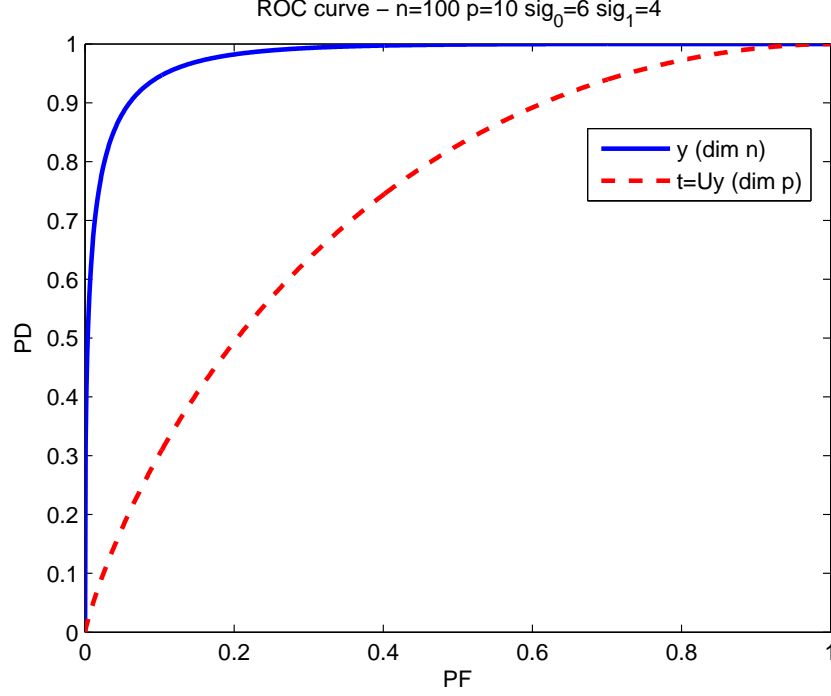
Stochastic Model

We begin with the following stochastic binary classification problem:

$$y = \begin{cases} U_0 x + z & y \in H_0 \\ U_1 x + z & y \in H_1 \end{cases} \quad (21)$$

where $z \in \mathbf{R}^n, z \sim \mathcal{N}(0, I)$. We assume that $U_0, U_1 \in \mathbf{R}^{n \times p}$ are known and have orthonormal columns. We assume $x \in \mathbf{R}^p$ and $f(x|H_0) \sim \mathcal{N}(0, \Sigma_0)$, $f(x|H_1) \sim \mathcal{N}(0, \Sigma_1)$. An extension to (2), shows that under $H_0, y \sim \mathcal{N}(0, U_0 \Sigma_0 U_0^T + I)$ and under $H_1, y \sim \mathcal{N}(0, U_1 \Sigma_1 U_1^T + I)$

We now consider the case where we are not given y explicitly but instead a processed version of it. That is



$$t = \begin{cases} U_0^T y & t \in H_0 \\ U_1^T y & t \in H_1 \end{cases} \quad (22)$$

By properties of Gaussian random variables, under H_0 , $y \sim \mathcal{N}(0, U_0^T U_0 \Sigma_0 U_0^T U_0 + U_0^T U_0)$ and under H_1 , $y \sim \mathcal{N}(0, U_1^T U_1 \Sigma_1 U_1^T U_1 + U_1^T U_1)$. Recalling that U_0, U_1 have orthonormal columns, we have that under H_0 , $t \sim \mathcal{N}(0, I + \Sigma_0)$ and under H_1 , $t \sim \mathcal{N}(0, I + \Sigma_1)$. Our LRT is

$$\begin{aligned} \Lambda(t) = \frac{f_1(t)}{f_0(t)} &= \frac{(2\pi)^{-p/2} |I + \Sigma_0|^{-1/2} \exp\{-\frac{1}{2} t^T (I + \Sigma_0)^{-1} t\}}{(2\pi)^{-p/2} |I + \Sigma_1|^{-1/2} \exp\{-\frac{1}{2} t^T (I + \Sigma_1)^{-1} t\}} \\ &= \left(\frac{|I + \Sigma_1|}{|I + \Sigma_0|} \right)^{1/2} \exp\left\{-\frac{1}{2} t^T \left[(I + \Sigma_0)^{-1} - (I + \Sigma_1)^{-1} \right] t\right\} \end{aligned} \quad (23)$$

where our decision is

$$\begin{aligned} &\text{Declare } H_0 \text{ if } \Lambda(t) < \eta \\ &\text{Declare } H_1 \text{ if } \Lambda(t) > \eta \end{aligned} \quad (24)$$

Since Σ_0, Σ_1 are valid covariance matrices, $|I + \Sigma_0| \geq 0$, $|I + \Sigma_1| \geq 0$. Therefore, if we define

$$\tilde{\Lambda}(t) = t^T \left[(I + \Sigma_0)^{-1} - (I + \Sigma_1)^{-1} \right] t \quad (25)$$

our decision becomes

$$\begin{aligned} &\text{Declare } H_0 \text{ if } \tilde{\Lambda}(t) < 2 \ln(\eta) + \ln \left(\frac{|I + \Sigma_0|}{|I + \Sigma_1|} \right) \\ &\text{Declare } H_1 \text{ if } \tilde{\Lambda}(t) > 2 \ln(\eta) + \ln \left(\frac{|I + \Sigma_0|}{|I + \Sigma_1|} \right) \end{aligned} \quad (26)$$

Defining $A = (I + \Sigma_0)^{-1} - (I + \Sigma_1)^{-1}$, our statistic is $\tau = t^T A t$.

There is no closed form solution for the distribution of τ unless A is diagonal, which results from Σ_1, Σ_0 also being diagonal and some other cases as well.

Stochastic setting Estimated Subspaces

We again consider

$$y = \begin{cases} U_0 x + z & y \in H_0 \\ U_1 x + z & y \in H_1 \end{cases} \quad (27)$$

where $z \in \mathbf{R}^n, z \sim \mathcal{N}(0, I)$. We assume that $U_0, U_1 \in \mathbf{R}^{n \times p}$ are unknown and have orthonormal columns. We assume $x \in \mathbf{R}^p$ and $f(x|H_0) \sim \mathcal{N}(0, \Sigma_0)$, $f(x|H_1) \sim \mathcal{N}(0, \Sigma_1)$. An extension to (2), shows that under H_0 , $y \sim \mathcal{N}(0, U_0 \Sigma_0 U_0^T + I)$ and under H_1 , $y \sim \mathcal{N}(0, U_1 \Sigma_1 U_1^T + I)$

We now consider the case where we are not given y explicitly but instead a processed version of it, using estimates \hat{U}_0, \hat{U}_1 of our underlying subspaces, U_0, U_1 . $U_0, U_1, \hat{U}_0, \hat{U}_1$ all have orthonormal columns. We distinguish between our testing and training data. Our estimates \hat{U}_0, \hat{U}_1 are formed from our training data:

$$y_{\text{train}} = \begin{cases} U_0 x_{\text{train}} + z_{\text{train}} & y_{\text{train}} \in H_0 \\ U_1 x_{\text{train}} + z_{\text{train}} & y_{\text{train}} \in H_1 \end{cases} \quad (28)$$

where $z_{\text{train}} \in \mathbf{R}^n, z \sim \mathcal{N}(0, I)$. We assume that $U_0, U_1 \in \mathbf{R}^{n \times p}$ are unknown and have orthonormal columns. We assume $x_{\text{train}} \in \mathbf{R}^p$ and $f(x_{\text{train}}|H_0) \sim \mathcal{N}(0, \Sigma_{\text{train}0})$, $f(x_{\text{train}}|H_1) \sim \mathcal{N}(0, \Sigma_{\text{train}1})$. Our estimates are formed from the leftmost left singular vectors of the SVD composition of the matrix formed by stacking the labeled training data in columns.

Our problem becomes

$$t = \begin{cases} \hat{U}_0^T y & t \in H_0 \\ \hat{U}_1^T y & t \in H_1 \end{cases} \quad (29)$$

By properties of Gaussian random variables, under H_0 , $y \sim \mathcal{N}(0, \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0 + \hat{U}_0^T \hat{U}_0)$ and under H_1 , $y \sim \mathcal{N}(0, \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1 + \hat{U}_1^T \hat{U}_1)$. Recalling that \hat{U}_0, \hat{U}_1 have orthonormal columns, we have that under H_0 , $t \sim \mathcal{N}(0, I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0)$ and under H_1 , $t \sim \mathcal{N}(0, I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1)$. Our LRT is

$$\begin{aligned} \Lambda(t) = \frac{f_1(t)}{f_0(t)} &= \frac{(2\pi)^{-p/2} |I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0|^{-1/2} \exp\{-\frac{1}{2} t^T (I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0)^{-1} t\}}{(2\pi)^{-p/2} |I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1|^{-1/2} \exp\{-\frac{1}{2} t^T (I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1)^{-1} t\}} \\ &= \left(\frac{|I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1|}{|I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0|} \right)^{1/2} \exp\left\{-\frac{1}{2} t^T \left[\left(I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0 \right)^{-1} - \left(I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1 \right)^{-1} \right] t\right\} \end{aligned} \quad (30)$$

where our decision is

$$\begin{aligned} &\text{Declare } H_0 \text{ if } \Lambda(t) < \eta \\ &\text{Declare } H_1 \text{ if } \Lambda(t) > \eta \end{aligned} \quad (31)$$

Defining

$$\tilde{\Lambda}(t) = t^T \left[\left(I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1 \right)^{-1} - \left(I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0 \right)^{-1} \right] t \quad (32)$$

We have the simplified decision

$$\begin{aligned}
& \text{Declare } H_0 \text{ if } \tilde{\Lambda}(t) < 2\ln(\eta) + \ln \left(\frac{|I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0|}{|I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1|} \right) \\
& \text{Declare } H_1 \text{ if } \tilde{\Lambda}(t) > 2\ln(\eta) + \ln \left(\frac{|I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0|}{|I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1|} \right)
\end{aligned} \tag{33}$$

To simplify this expression we turn to random matrix theory. We assume that our subspace estimate \hat{U}_0 is formed from n training samples. We consider the matrix $\hat{U}_0 U_0$ whose $i^{\text{th}}, j^{\text{th}}$ entry is $\langle \hat{u}_{0i}, u_{0j} \rangle$. From random matrix theory we know that for $j \neq i$, $\langle \hat{u}_{0i}, u_{0j} \rangle \approx \frac{1}{n} \approx 0$ for large enough n . For $i = j$,

$$\langle \hat{u}_{0j}, u_{0j} \rangle \approx \sqrt{\frac{\Sigma_{\text{train}0jj}^2 - \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}}}$$

Assuming that $\Sigma_{\text{train}0}$ is diagonal we have that

$$\begin{aligned}
\hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0 & \approx \text{diag} \left(\sqrt{\frac{\Sigma_{\text{train}0jj}^2 - \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}}} \right) \text{diag}(\Sigma_{0jj}) \text{diag} \left(\sqrt{\frac{\Sigma_{\text{train}0jj}^2 - \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}}} \right) \\
& = \text{diag} \left(\Sigma_{0jj} \left(\frac{\Sigma_{\text{train}0jj}^2 - \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}} \right) \right)
\end{aligned} \tag{34}$$

Using this expression, we may simplify (32) and (33).

$$\begin{aligned}
& \left(I + \hat{U}_0^T U_0 \Sigma_0 U_0^T \hat{U}_0 \right)^{-1} - \left(I + \hat{U}_1^T U_1 \Sigma_1 U_1^T \hat{U}_1 \right)^{-1} \\
& \approx \left(I + \text{diag} \left(\Sigma_{0jj} \left(\frac{\Sigma_{\text{train}0jj}^2 - \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}} \right) \right) \right)^{-1} - \left(I + \text{diag} \left(\Sigma_{1jj} \left(\frac{\Sigma_{\text{train}1jj}^2 - \frac{p}{n}}{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n}} \right) \right) \right)^{-1} \\
& = \left(\text{diag} \left(\frac{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n} + \Sigma_{0jj} (\Sigma_{\text{train}0jj}^2 - \frac{p}{n})}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}} \right) \right)^{-1} - \left(\text{diag} \left(\frac{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n} + \Sigma_{1jj} (\Sigma_{\text{train}1jj}^2 - \frac{p}{n})}{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n}} \right) \right)^{-1} \\
& = \text{diag} \left(\frac{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n} + \Sigma_{0jj} (\Sigma_{\text{train}0jj}^2 - \frac{p}{n})} \right) - \text{diag} \left(\frac{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n}}{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n} + \Sigma_{1jj} (\Sigma_{\text{train}1jj}^2 - \frac{p}{n})} \right) \\
& := D
\end{aligned} \tag{35}$$

Define

$$\begin{aligned}
\hat{\Lambda}(t) & \approx t^T D t = y^T \hat{U}_i D \hat{U}_i^T y \\
& = \sum_{j=1}^p t_j^2 \left(\frac{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n}}{\Sigma_{\text{train}0jj}^2 + \Sigma_{\text{train}0jj} \frac{p}{n} + \Sigma_{0jj} (\Sigma_{\text{train}0jj}^2 - \frac{p}{n})} - \frac{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n}}{\Sigma_{\text{train}1jj}^2 + \Sigma_{\text{train}1jj} \frac{p}{n} + \Sigma_{1jj} (\Sigma_{\text{train}1jj}^2 - \frac{p}{n})} \right)
\end{aligned} \tag{36}$$

Then our decision becomes (assuming that $\eta = 1$)

$$\begin{aligned}
\text{Declare } H_0 \text{ if } \hat{\Lambda}(t) &< \sum_{j=1}^p \ln \left(\frac{\left(\Sigma_{\text{train}1_{jj}}^2 + \Sigma_{\text{train}1_{jj}} \frac{p}{n} \right) \left(\Sigma_{\text{train}0_{jj}}^2 + \Sigma_{\text{train}0_{jj}} \frac{p}{n} + \Sigma_{0_{jj}} (\Sigma_{\text{train}0_{jj}}^2 - \frac{p}{n}) \right)}{\left(\Sigma_{\text{train}0_{jj}}^2 + \Sigma_{\text{train}0_{jj}} \frac{p}{n} \right) \left(\Sigma_{\text{train}1_{jj}}^2 + \Sigma_{\text{train}1_{jj}} \frac{p}{n} + \Sigma_{1_{jj}} (\Sigma_{\text{train}1_{jj}}^2 - \frac{p}{n}) \right)} \right) \\
\text{Declare } H_1 \text{ if } \hat{\Lambda}(t) &> \sum_{j=1}^p \ln \left(\frac{\left(\Sigma_{\text{train}1_{jj}}^2 + \Sigma_{\text{train}1_{jj}} \frac{p}{n} \right) \left(\Sigma_{\text{train}0_{jj}}^2 + \Sigma_{\text{train}0_{jj}} \frac{p}{n} + \Sigma_{0_{jj}} (\Sigma_{\text{train}0_{jj}}^2 - \frac{p}{n}) \right)}{\left(\Sigma_{\text{train}0_{jj}}^2 + \Sigma_{\text{train}0_{jj}} \frac{p}{n} \right) \left(\Sigma_{\text{train}1_{jj}}^2 + \Sigma_{\text{train}1_{jj}} \frac{p}{n} + \Sigma_{1_{jj}} (\Sigma_{\text{train}1_{jj}}^2 - \frac{p}{n}) \right)} \right)
\end{aligned} \tag{38}$$