11. HAFTA

Varyans-Kovaryans Matrisleri Eşit Olmayan İki Çok Değişkenli Normal Kitle için Sınıflandırma

$\Sigma_1 \neq \Sigma_2$ **olsun**

Kitle varyans-kovaryans matrisleri eşit olmadığında, sınıflandırma kuralları daha karmaşıktır. $\Sigma_1 = \Sigma_2$ olduğu durumdaki benzer işlemler yapıldığında, sınıflandırma bölgeleri

$$R_1 : -\frac{1}{2}\underline{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\underline{x} + (\underline{\mu}_1'\Sigma_1^{-1} - \underline{\mu}_2'\Sigma_2^{-1})\underline{x} - k \geq \ln\left(\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{p_2}{p_1}\right] \right)$$

$$R_2 : -\frac{1}{2}\underline{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\underline{x} + (\underline{\mu}_1'\Sigma_1^{-1} - \underline{\mu}_2'\Sigma_2^{-1})\underline{x} - k < \ln\left(\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{p_2}{p_1}\right] \right)$$

dır, burada

$$k = \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\underline{\mu}_1'\Sigma_1^{-1}\underline{\mu}_1 - \underline{\mu}_2'\Sigma_2^{-1}\underline{\mu}_2)$$

dır.

Sınıflandırma bölgeleri $\underline{x}$'in karesel fonksiyonları ile tanımlandığından, elde edilen bu fonksiyona karesel diskriminant fonksiyonu adı verilir. Karesel ifade $-\frac{1}{2}\underline{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\underline{x}$ dir.

Sonuç : Π_1 ve Π_2 , yoğunluk fonksiyonları $f_1(\underline{x})$ ve $f_2(\underline{x})$, kitle ortalama vektörleri $\underline{\mu}_1$, $\underline{\mu}_2$ ve varyans–kovaryans matrisleri Σ_1, Σ_2 olan çok değişkenli (p-boyutlu) normal dağılıma sahip kitleler olduğunu kabul edelim. Bu durumda hatalı sınıflandırmanın beklenen maliyetini minimize eden atama kuralı

$$-\frac{1}{2}\underline{x}_0'(\Sigma_1^{-1} - \Sigma_2^{-1})\underline{x}_0 + (\underline{\mu}_1'\Sigma_1^{-1} - \underline{\mu}_2'\Sigma_2^{-1})\underline{x}_0 - k \geq \ln\left(\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{p_2}{p_1}\right] \right)$$

ise $\underline{x}_0$ gözlem değerine sahip birim Π_1'e aksi halde Π_2'ye atanır biçiminde verilir.

Eğer parametreler bilinmiyor ise, örneklemlerden elde edilen tahminlere bağlı örneklem sınıflandırma kuralı,

$$-\frac{1}{2}\underline{x}_0'(S_1^{-1}-S_2^{-1})\underline{x}_0+(\overline{\underline{x}}_1{}'S_1^{-1}-\overline{\underline{x}}_2{}'S_2^{-1})\underline{x}_0-\hat{k}\geq \ln(\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{p_2}{p_1}\right])$$

ise $\underline{x}_0$ gözlem değerine sahip birim Π_1'e, aksi halde Π_2'ye atanır biçiminde verilir. Burada,

$$\hat{k}=\frac{1}{2}\ln(\frac{|S_1|}{|S_2|})+\frac{1}{2}(\overline{\underline{x}}_1{}'\Sigma_1^{-1}\overline{\underline{x}}_1-\overline{\underline{x}}_2{}'\Sigma_2^{-1}\overline{\underline{x}}_2)$$

dir.

Rasgele vektörlerin dağılımı her zaman çok değişkenli normal olmak zorunda değildir. Normale dönüşüm yapılabileceği gibi, bu dönüşüm yapılmadan da normal olmayan kitleler için sınıflandırma yapılır.

Sınıflandırma (Atama) Kurallarının Değerlendirilmesi

Herhangi bir sınıflandırma yönteminin performansına karar vermenin önemli bir yolu, hatalı sınıflandırma olasılıkları veya hata oranlarına göre karar vermektir. Rasgele değişkenlerin dağılımı veya kitleler tamamen bilindiğinde, hatalı sınıflandırma olasılıkları kolay hesaplanabilir. Kitleler tamamen bilinmediğinde (yani bazı parametreler bilinmiyor olabilir) örneklem sınıflandırma fonksiyonları üzerinde durulur. Öncelikle sınıflandırma fonksiyonu oluşturulur ve fonksiyonun performansı gerçek örneklemlerle değerlendirilir.

İki kitle için maliyetlerin eşit olduğu durumda elde edilen ECM sınıflandırma kuralına karşılık gelen Toplam Hatalı Sınıflandırma Olasılığının (*TPM*),

$$TPM=p_1\int_{R_2}f_1(\underline{x})d\underline{x}+p_2\int_{R_1}f_2(\underline{x})d\underline{x}$$

olarak verilmişti. Seçilen R_1 ve R_2 bölgelerine göre elde edilen TPM nin enküçük değerine Optimal (En iyi) Hata Oranı (OER) adı verilir. OER, minimum TPM sınıflandırma kuralı için bir hata oranıdır.

Örnek 22 : $p_1 = p_2 = 1$ ve $f_1(\underline{x})$ ve $f_2(\underline{x})$ ortalama vektörleri $\underline{\mu}_1$, $\underline{\mu}_2$ ve varyans–kovaryans matrisleri $\Sigma_1 = \Sigma_1 = \Sigma$ olan çok değişkenli (*p*-boyutlu) normal dağılıma sahip kitleler olduğunu, optimal hata oranı için ifadeler elde edilsin.

Minimum ECM ve minimum TPM sınıflandırma kuralları, $C(1/2) = C(2/1)$ olduğunda birbirine eşdeğer iki kuraldır. Önsel olasılıklar da eşit olduğunda, $\ln\left[\left(\dfrac{C(1/2)}{C(2/1)}\right)\left(\dfrac{p_2}{p_1}\right)\right] = 0$ ile minimum TPM sınıflandırma bölgeleri tanımlanabilir. Buradan sınıflandırma bölgeleri;

$$R_1 : (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{x} - \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2) \geq 0$$

ve

$$R_2 : (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{x} - \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2) < 0$$

biçiminde bulunur. Bu kümeler $y = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{x} = \underline{l}'\underline{x}$ ile tanımlanır. Burada

$$R_1(\mathrm{y}) : y \geq \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2)$$

ve

$$R_2(\mathrm{y}) : y < \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2)$$

dir. *Y* rasgele değişkeni normal rasgele değişkenlerin lineer bir birleşimi olduğundan, *Y*'nin olasılık yoğunluk fonksiyonu $f_1(\mathrm{y})$ ve $f_2(\mathrm{y})$ tek değişkenli normaldir. Burada

$$Y = \underline{l}'\underline{X}$$

olmak üzere

$$\begin{aligned} E(Y/\Pi_1) &= E(\underline{l}'\underline{X}/\Pi_1) \\ &= \underline{l}'E(\underline{X}/\Pi_1) \\ &= \underline{l}'\underline{\mu}_1 \qquad , \\ &= (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{\mu}_1 \\ &= \mu_{1Y} \end{aligned}$$

$$
\begin{aligned}
E(Y/\Pi_2) &= E(\underline{l}'\underline{X}/\Pi_2)\\
&= \underline{l}'E(\underline{X}/\Pi_2)\\
&= \underline{l}'\underline{\mu}_2\\
&= (\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{\mu}_2\\
&= \mu_{2Y}
\end{aligned}
$$

ve

$$
\begin{aligned}
Var(Y/\Pi_1) &= Var(Y/\Pi_2)\\
&= Var(Y)\\
&= Var(\underline{l}'\underline{X})\\
&= \underline{l}'Cov(\underline{X})\underline{l}\\
&= \underline{l}'\Sigma\underline{l}\\
&= (\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2)\\
&= \Delta^2\\
&= \sigma_Y^2
\end{aligned}
$$

biçiminde elde edilir. Burada, Δ^2 iki kitle arasındaki Mahalanobis kare uzaklığıdır. Buradan,

$$
\begin{aligned}
TPM = &\frac{1}{2}P(\Pi_1 \text{ den bir gözlemin } \Pi_2 \text{ ye hatalı sınıflandırılması})\\
&+\frac{1}{2}P(\Pi_2 \text{ den bir gözlemin } \Pi_1 \text{ ye hatalı sınıflandırılması})
\end{aligned}
$$

dır.

$$
\begin{aligned}
P(\Pi_1 \text{ den bir gözlemin } \Pi_2 \text{ ye hatalı sınıflandırılması}) &= P(\underline{X}\in R_2/\Pi_1)\\
&= P(2/1)\\
&= P(Y<\frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)/\Pi_1)\\
&= P(\frac{Y-\mu_{1Y}}{\sigma_Y}<\frac{\frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)-\mu_{1Y}}{\sigma_Y})\\
&= P(Z<\frac{\frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)-(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{\mu}_1}{\Delta})\\
&= P(Z<\frac{-\frac{1}{2}\Delta^2}{\Delta})\\
&= P(Z<-\frac{\Delta}{2})\\
&= \Phi\left\{-\frac{\Delta}{2}\right\}
\end{aligned}
$$

dir, burada $\Phi\{.\}$ standart normal bir rasgele değişkenin birikimli dağılım fonksiyonunu göstermektedir. Benzer biçimde,

$$
\begin{aligned}
P(\Pi_2 \text{ den bir gözlemin } \Pi_1 \text{ e hatalı sınıflandırılması}) &= P(\underline{X} \in R_1 / \Pi_2) \\
&= P(1/2) \\
&= P(Y \geq \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2) / \Pi_2) \\
&= P(\frac{Y - \mu_{2Y}}{\sigma_Y} \geq \frac{\frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2) - \mu_{2Y}}{\sigma_Y}) \\
&= P(Z \geq \frac{\frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2) - (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{\mu}_2}{\Delta}) \\
&= P(Z \geq \frac{\frac{1}{2}\Delta^2}{\Delta}) \\
&= P(Z \geq \frac{\Delta}{2}) \\
&= 1 - P(Z < \frac{\Delta}{2}) \\
&= 1 - \Phi\left\{\frac{\Delta}{2}\right\} \\
&= \Phi\left\{-\frac{\Delta}{2}\right\}
\end{aligned}
$$

dir. Böylece Optimal Hata Oranı

$$
\begin{aligned}
OER = Minimum\ TPM &= \frac{1}{2}\Phi\left\{-\frac{\Delta}{2}\right\} + \frac{1}{2}\Phi\left\{-\frac{\Delta}{2}\right\} \\
&= \Phi\left\{-\frac{\Delta}{2}\right\}
\end{aligned}
$$

olarak elde edilir. Örneğin,

$\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = 2.56$ alınırsa, $\Delta = \sqrt{2.56} = 1.6$ olur ve

$$
\begin{aligned}
OER = Minimum\ TPM &= \frac{1}{2}\Phi\left\{-\frac{1.6}{2}\right\} + \frac{1}{2}\Phi\left\{-\frac{1.6}{2}\right\} \\
&= \Phi\left\{-\frac{1.6}{2}\right\} \\
&= 0.2119
\end{aligned}
$$

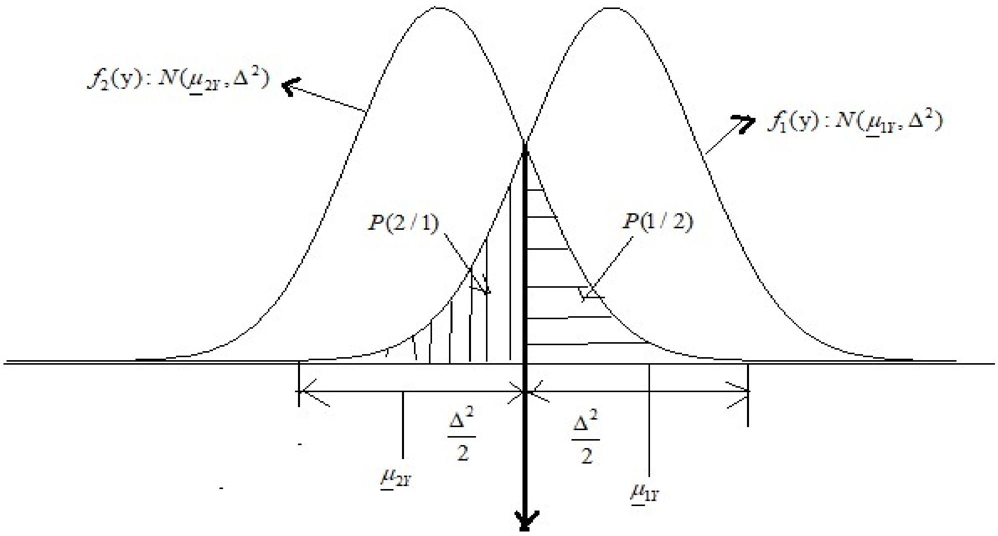
olarak bulunur. Buradan birimlerin %21.19 'nun hatalı sınıflandırıldığı veya diğer bir ifade ile hata oranın %21.19 olduğu görülmektedir.

$$\mu_{1Y} - \text{orta nokta} = \underbrace{\left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)^{'} \Sigma^{-1} \underset{\sim}{\mu}_1}_{\mu_{1Y}} - \underbrace{\frac{1}{2}\left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)^{'} \Sigma^{-1} \left(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2\right)}_{\text{orta nokta}} = \frac{\Delta^2}{2}$$

ve

$$\mu_{2Y} - \text{orta nokta} = \underbrace{\left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)^{'} \Sigma^{-1} \underset{\sim}{\mu}_2}_{\mu_{1Y}} - \underbrace{\frac{1}{2}\left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)^{'} \Sigma^{-1} \left(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2\right)}_{\text{orta nokta}} = \frac{\Delta^2}{2}$$

dir.



$$Orta\ Nokta = \frac{1}{2}\left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)^{'} \Sigma^{-1} \left(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2\right)$$

Bu örnekte kitle yoğunluk fonksiyonları tamam bilindiğinde yani kitle parametrelerinin de bilindiği durumda optimal hata oranın nasıl hesaplanacağı gösterilmiştir. Ancak kitle parametreleri bilinmediğinde örneklemlerden elde edilen tahmin değerleri alındığında, hata oranları kolay hesaplanamaz.

Örneklem sınıflandırma fonksiyonlarının performansı Gerçek Hata Oranı (AER) nin hesaplanmasıyla görülebilir. Bu hata oranına koşullu hata oranı da denmektedir. Gerçek hata oranı

$$AER = p_1 \int_{\hat{R}_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{\hat{R}_1} f_2(\underline{x}) d\underline{x}$$

ifadesinden elde edilir. Burada $\hat{R}_1$ ve $\hat{R}_2$, n_1 ve n_2 birimlik örneklemlerden belirlenen sınıflandırma bölgelerini göstermektedir.

$$\begin{aligned}W &= (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{x}_0 - \frac{1}{2}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2) \\ &= (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1}\left[\underline{x}_0 - \frac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)\right]\end{aligned}$$

sınıflandırma fonksiyonu söz konusu olduğunda, $\hat{R}_1$ ve $\hat{R}_2$ bölgeleri aşağıdaki eşitsizlikleri sağlayan $\underline{x}$ lerin kümesi ile belirlenir. Bu bölgeler,

$$\hat{R}_1 : (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{x} - \frac{1}{2}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2) \geq \ln\left(\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{p_2}{p_1}\right] \right)$$

ve

$$\hat{R}_2 : (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{x} - \frac{1}{2}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2) < \ln\left(\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{p_2}{p_1}\right] \right)$$

olarak belirlenir.

Diğer bir hata oranı, gerçek hata oranının beklenen değeri olan beklenen gerçek hata oranıdır. Bu hata oranına koşulsuz hata oranı da denmektedir ve $E(AER)$ biçiminde gösterilir. Beklenen gerçek(koşulsuz) hata oranı, gerçek (koşullu) hata oranının tüm örneklemler üzerinden beklenen değeri alınarak elde edilir. *AER* parametrelerin örneklemlerden elde edilen tahmin edicilerine bağlı bir hata oranı olduğundan, rasgele bir değişkendir.

Gerçek hata oranı *AER* örneklem sınıflandırma fonksiyonunun gelecek örneklemlerdeki performansının nasıl olacağını verir. Gerçek hata oranı bilinmeyen $f_1(\underline{x})$ ve $f_2(\underline{x})$ yoğunluk fonksiyonlarına bağlı olduğundan, optimal hata oranına benzer biçimde hesaplanamaz. Ancak gerçek hata oranıyla ilişkili bir değerin tahmini elde edilebilir.

Sınıflandırma fonksiyonunun performansını ölçen öyle ölçütler vardır ki kitlelerin dağılımlarına bağlı değildir ve her sınıflandırma yöntemi için hesaplanabilir. Görünüşte hata oranı (*APER*) adı verilen bu ölçü, örneklem sınıflandırma fonksiyonu ile hatalı sınıflandırılan deneme örneklemindeki gözlem sayısının, tüm örneklem sayısına oranı biçiminde tanımlanır. Π_1'den alınan n_1 gözlemden, n_{11} tanesi Π_1'e ve n_{12} tanesi Π_2'ye ve Π_2'den alınan n_2 gözlemden, n_{22} tanesi Π_2'ye ve n_{21} tanesi Π_1'e atandığında görünüşte hata oranı

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2}$$

biçiminde elde edilir.

Görünüşte hata oranı kolay elde edilebilir ve dağılım varsayımı gerektirmez. Bu hata oranının elde edilmesine ilişkin farklı yaklaşımlar da vardır. Örneğin yeterince büyük örneklem var ise, gözlemlerin yarısıyla diskriminant fonksiyonu oluşturulur, geriye kalan diğer gözlemlerle de bu fonksiyonun performansı değerlendirilir.

İkiden çok kitle olduğunda da benzer yöntemler uygulanarak sınıflandırma yapılabilir.