Derivation of Linear Discriminant Analysis Formula

 만약 g개의 그룹이 있다면, 베이즈 법칙은 $P(i\,|\,x) > P(j\,|\,x),\ \forall\, i \neq j$에서 가장 큰 조건부 확률을 가지는 개체를 그룹 i에 할당함으로써 분류의 총 오류를 최소화 한다. $P(i\,|\,x)$ (즉, 측정값이 주어졌을 때 해당 클래스의 확률이 무엇인지)를 측정값으로부터 직접적으로 구할 수 없고 $P(x\,|\,i)$ (즉, 클래스가 주어졌을 때 측정값을 알면 각각의 클래스에 대한 확률을 계산할 수 있다.)를 구할 수 있기 때문에 베이즈 정리를 사용 한다:

$$P(i\,|\,x) = \frac{P(x\,|\,i)\,P(i)}{\sum_{all\,j} P(x|j)P(j)}$$

그러므로 베이즈 법칙은 $\dfrac{P(x\,|\,i)\,P(i)}{\sum_{all\,k} P(x|k)P(k)} > \dfrac{P(x\,|\,j)\,P(j)}{\sum_{all\,k} P(x|k)P(k)}$, $\forall\ j \neq i$을 만족하면 개체를 그룹 i에 할당.

부등식의 양변의 분모는 양수이고 같은 값을 갖는다. 그러므로 분모를 소거할 수 있다.

$$P(x\,|\,i)\,P(i) > P(x\,|\,j)\,P(j)\ ,\ \forall\ j \neq i$$

만약 우리가 많은 클래스와 각각의 차원이 많은 값을 가질 측정값의 여러 차원을 가지고 있다면, 조건부 확률 $P(x\,|\,i)$의 계산은 많은 데이터를 필요로 한다. 데이터가 어떤 이론적인 분포로부터 온다고 가정하는 것이 더 실용적이다. 가장 널리 쓰이는 가정은 우리의 데이터가 다변량 정규분포로부터 왔다는 것이다.

$$P(x\,|\,i) = \left(\frac{1}{(2\pi)^{\frac{n}{2}}\,|\,C_i|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right)$$

μ_i는 mean vector이고, C_i는 그룹 i의 covariance matrix이다.

이 분포 formula를 베이즈 법칙에 대입하면

측정값이 x인 개체를 그룹 i에 할당한다. if

$$\left(\frac{P(i)}{(2\pi)^{\frac{n}{2}}\,|\,C_i|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right) > \left(\frac{P(j)}{(2\pi)^{\frac{n}{2}}\,|\,C_j|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_j)^T C_j^{-1}(x-\mu_j)\right), i \neq j$$

$(2\pi)^{\frac{n}{2}}$가 양변에 같이 있기 때문에 소거하면

$$\left(\frac{P(i)}{|\,C_i|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right) > \left(\frac{P(j)}{|\,C_j|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_j)^T C_j^{-1}(x-\mu_j)\right),\ i \neq j$$

양변에 자연로그를 취하면

$$-\frac{1}{2}ln(|C_i|)+\ln(P(i))-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i) > -\frac{1}{2}ln(|C_j|)+\ln(P(j))-\frac{1}{2}(x-\mu_j)^T C_j^{-1}(x-\mu_j)$$

양변에 –2를 곱함

$$\ln(|C_i|)-2\ln(P(i))+(x-\mu_i)^T C_i^{-1}(x-\mu_i) < \ln(|C_j|)-2\ln(P(j))+(x-\mu_j)^T C_j^{-1}(x-\mu_j),\ i \neq j$$

$d_i(x) = \ln(|C_i|)+(x-\mu_i)^T C_i^{-1}(x-\mu_i)$라 하면,

측정값이 x인 개체를 그룹 i에 할당한다. if

$$d_i(x)-2\ln(P(i)) < d_j(x)-2\ln(P(j)), i \neq j$$

That is Quadratic Discriminant function.

만약 모든 covariance matrix가 $C = C_i = C_j$로 동일하다면

$$\ln(|C|)-2\ln(P(i))+(x-\mu_i)^T C^{-1}(x-\mu_i) < \ln(|C|)-2\ln(P(j))+(x-\mu_j)^T C^{-1}(x-\mu_j),\ i \neq j$$

$(x-\mu_i)^T C^{-1}(x-\mu_i)$를 $xC^{-1}x^T-2\mu_i C^{-1}x^T+\mu_i C^{-1}\mu_i^T$로 쓸 수 있으므로

$$\ln(|C|)-2\ln(P(i))+xC^{-1}x^T-2\mu_i C^{-1}x^T+\mu_i C^{-1}\mu_i^T <$$
$$\ln(|C|)-2\ln(P(j))+xC^{-1}x^T-2\mu_j C^{-1}x^T+\mu_j C^{-1}\mu_j^T,\ i \neq j$$

$\ln(|C|)$와 $xC^{-1}x^T$소거

$$-2\ln(P(i))-2\mu_i C^{-1}x^T+\mu_i C^{-1}\mu_i^T\ <\ -2\ln(P(j))-2\mu_j C^{-1}x^T+\mu_j C^{-1}\mu_j^T,\ i \neq j$$

$-\frac{1}{2}$을 양변에 곱하면

$$\ln(P(i))+\mu_i C^{-1}x^T-\frac{1}{2}\mu_i C^{-1}\mu_i^T\ >\ \ln(P(j))+\mu_j C^{-1}x^T-\frac{1}{2}\mu_j C^{-1}\mu_j^T,\ i \neq j$$

$f_i = \ln(P(i))+\mu_i C^{-1}x^T-\frac{1}{2}\mu_i C^{-1}\mu_i^T$로 놓으면,

측정값이 x인 개체를 그룹 i에 할당한다. if $f_i > f_j,\ \forall\ i \neq j$

That is Linear Discriminant function.