4.16

Intro

In this exercise we will perform the simplification of likelihood ratios for Gaussians. We will study four cases, going from the most general (Σ matrices are different and each have $\frac{d(d+1)}{2}$) to the most simple (Σ matrices are equal and have only one parameter σ^2). Since the classifier is binary, we have $K = 2$. Moreover, I should point out that I made all the simplifications that I found relevant. If you think there is more simplifications that are relevant, please post on the issues of the GitHub page :).

Solution

Form of Σ_j: Arbitraty (Σ_j)

The likelihood can be expressed as:

$$\begin{aligned}
&\frac{p(x|y=1)}{p(x|y=0)} = \frac{\mathcal{N}(x|\mu_1, \Sigma_1)}{\mathcal{N}(x|\mu_0, \Sigma_0)} = \\
&\frac{|\Sigma_1^{-\frac{1}{2}}|}{|\Sigma_0^{-\frac{1}{2}}|} exp\left[-\frac{1}{2}\left((x-\mu_1)^t\Sigma_1^{-1}(x-\mu_1) - (x-\mu_0)^t\Sigma_0^{-1}(x-\mu_0)\right)\right]
\end{aligned} \tag{1}$$

where the ony simplification is the elimination of the term $(2\pi)^{-\frac{D}{2}}$. Taking the log of the Equation 1, we arrive at:

$$\begin{aligned}
&log\frac{p(x|y=1)}{p(x|y=0)} = \\
&-\frac{1}{2}\left[log\frac{|\Sigma_1|}{|\Sigma_0|} + (x-\mu_1)^t\Sigma_1^{-1}(x-\mu_1) - (x-\mu_0)^t\Sigma_0^{-1}(x-\mu_0)\right]
\end{aligned} \tag{2}$$

Form of Σ_j: Shared ($\Sigma_j = \Sigma$)

When the two covariance matrices are equal, Equation (2) can be simplified to

$$\begin{aligned}
&-\frac{1}{2}\left[log\frac{|\Sigma|}{|\Sigma|} + (x-\mu_1)^t\Sigma^{-1}(x-\mu_1) - (x-\mu_0)^t\Sigma^{-1}(x-\mu_0)\right] = \\
&-\frac{1}{2}\left[tr[(x-\mu_1)^t\Sigma^{-1}(x-\mu_1)] - tr[(x-\mu_2)^t\Sigma^{-1}(x-\mu_2)]\right] = \\
&-\frac{1}{2}\left[tr[\Sigma^{-1}[(x-\mu_1)(x-\mu_1)^t - (x-\mu_2)(x-\mu_2)^t]]\right]
\end{aligned} \tag{3}$$

Form of Σ_j: Shared, axis-aligned ($\Sigma_j = \Sigma$ with $\Sigma_{ij} = 0$ for $i \neq j$)

When the axis are aligned Equation (3) becomes:

$$\begin{aligned}
&-\frac{1}{2}\left[tr[\text{diag}(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, ...)[(x-\mu_1)(x-\mu_1)^t - (x-\mu_2)(x-\mu_2)^t]]\right] = \\
&-\frac{1}{2}\left[\sum_{i=1}^{d}\frac{1}{\sigma_i^2}[(x_i-\mu_{1,i})^2 - (x_i-\mu_{0,i})^2]\right] = \\
&-\frac{1}{2}\left[\sum_{i=1}^{d}\frac{1}{\sigma_i^2}(2x_i-\mu_{1,i}-\mu_{0,i})(\mu_{0,i}-\mu_{1,i})\right]
\end{aligned} \tag{4}$$

Form of Σ_j: Shared, spherical ($\Sigma_j = \sigma^2 I$)

As the last case, when we have spherical covariance matrices, Equation (4) becomes:

$$
\begin{aligned}
&-\frac{1}{2}\left[\sum_{i=1}^{d} \frac{1}{\sigma^2}(2x_i - \mu_{1,i} - \mu_{0,i})(\mu_{0,i} - \mu_{1,i})\right] = \\
&-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{d} (2x_i - \mu_{1,i} - \mu_{0,i})(\mu_{0,i} - \mu_{1,i})\right]
\end{aligned}
\tag{5}
$$

Conclusion

In this exercise we performed the simplification of the log likelihood ratio for Gaussians. We analysed four cases. The first was the most general, where the covariance matrices where not shared and had all parameters. In this cases the only simplification made was removing the normalization constant $(2\pi)^{-\frac{D}{2}}$. In the second case, the shared matrix property allowed to eliminate the log term all together and to merge the quadratic term expressions using the trace operator. In the third case, the use of the shared, axis-aligned matrices allowd to simplify the trace operation to a simple scalar expression. At last, in the final case, the spherical covariance matrices allowed to put the only covariance term in evidence.