Task 4 - Maximum Likelihood Estimator

In statistics, the Bayesian estimator minimizes the posterior expected value of a loss function. In the case of binary classification, this means that a Bayes estimator considers the class with a higher posteriori probability, e.g. if

$$p(\Delta|x) > p(\circ|x) \tag{1}$$

a Bayes estimator would assign class Δ.
For LDA, we assume Gaussian distributed data, s.t.

$$p(x|\Delta) = \frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|S_\Delta|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{w}_\Delta)^\top S_\Delta^{-1}(\mathbf{x}-\mathbf{w}_\Delta)}$$

Compute the Bayes classifier that fulfills (1), remember that

$$\begin{aligned}
& p(\Delta|x) > p(\circ|x) \\
\Leftrightarrow & \frac{p(\Delta|x)}{p(\circ|x)} > 1 \\
\Leftrightarrow & \log\left(\frac{p(\Delta|x)}{p(\circ|x)}\right) > 0.
\end{aligned}$$

Check slides 23 and 24 of Lecture 3 for further information.

Solution:

$$\begin{aligned}
&\log\left(\frac{p(\Delta|x)}{p(\circ|x)}\right) \\
=&\log(p(\Delta|x)) - \log(p(\circ|x)) \\
=&\log\left(\frac{p(x|\Delta)\cdot p(\Delta)}{p(x)}\right) - \log\left(\frac{p(x|\circ)\cdot p(\circ)}{p(x)}\right) \\
=&\log(p(x|\Delta)) + \log(p(\Delta)) - \cancel{\log(p(x))} - (\log(p(x|\circ)) + \log(p(\circ)) - \cancel{\log(p(x))}) \\
=&\log\left(\frac{p(\Delta)}{p(\circ)}\right) + \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|S_\Delta|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{w}_\Delta)^\top S_\Delta^{-1}(\mathbf{x}-\mathbf{w}_\Delta)}\right) - \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|S_\circ|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{w}_\circ)^\top S_\circ^{-1}(\mathbf{x}-\mathbf{w}_\circ)}\right) \\
&\text{we assume } S_\circ = S_\Delta = S_W \\
=&\log\left(\frac{p(\Delta)}{p(\circ)}\right) - \frac{1}{2}(\mathbf{x}-\mathbf{w}_\Delta)^\top S_W^{-1}(\mathbf{x}-\mathbf{w}_\Delta) + \frac{1}{2}(\mathbf{x}-\mathbf{w}_\circ)^\top S_W^{-1}(\mathbf{x}-\mathbf{w}_\circ) \\
=&\log\left(\frac{p(\Delta)}{p(\circ)}\right) + \frac{1}{2}(\mathbf{w}_\Delta^\top S_W^{-1}(\mathbf{x}-\mathbf{w}_\Delta) - \mathbf{x}^\top S_W^{-1}(\mathbf{x}-\mathbf{w}_\Delta) - \mathbf{w}_\circ^\top S_W^{-1}(\mathbf{x}-\mathbf{w}_\circ) + \mathbf{x}^\top S_W^{-1}(\mathbf{x}-\mathbf{w}_\circ)) \\
=&\log\left(\frac{p(\Delta)}{p(\circ)}\right) + \frac{1}{2}(\mathbf{w}_\Delta^\top S_W^{-1}\mathbf{x} - \mathbf{w}_\Delta^\top S_W^{-1}\mathbf{w}_\Delta - \cancel{\mathbf{x}^\top S_W^{-1}\mathbf{x}} + \mathbf{x}^\top S_W^{-1}\mathbf{w}_\Delta \\
&\qquad - \mathbf{w}_\circ^\top S_W^{-1}\mathbf{x} + \mathbf{w}_\circ^\top S_W^{-1}\mathbf{w}_\circ + \cancel{\mathbf{x}^\top S_W^{-1}\mathbf{x}} - \mathbf{x}^\top S_W^{-1}\mathbf{w}_\circ) \\
=&\log\left(\frac{p(\Delta)}{p(\circ)}\right) + \mathbf{w}_\Delta^\top S_W^{-1}\mathbf{x} - \mathbf{w}_\circ^\top S_W^{-1}\mathbf{x} - \frac{1}{2}\mathbf{w}_\Delta^\top S_W^{-1}\mathbf{w}_\Delta + \frac{1}{2}\mathbf{w}_\circ^\top S_W^{-1}\mathbf{w}_\circ \\
=&\underbrace{(\mathbf{w}_\Delta - \mathbf{w}_\circ)^\top S_W^{-1}}_{\mathbf{w}^\top}\mathbf{x} + \underbrace{\frac{1}{2}\left(\mathbf{w}_\circ^\top S_W^{-1}\mathbf{w}_\circ - \mathbf{w}_\Delta^\top S_W^{-1}\mathbf{w}_\Delta\right) + \log\left(\frac{p(\Delta)}{p(\circ)}\right)}_{\beta}
\end{aligned}$$

Note that if we want to compute the classification boundary, we are interested in (1) to be an equality.