Solution for the Exercise on Page 26 (Thursday Session)

Example 1. Consider the dissimilarity matrix

$$D = \begin{bmatrix} 0 & 1 & 1/4 & \infty & \infty & 1/4 \\ 1 & 0 & 1/4 & \infty & \infty & 1/4 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 & \infty \\ \infty & \infty & 1/4 & 0 & 1 & 1/4 \\ \infty & \infty & 1/4 & 1 & 0 & 1/4 \\ 1/4 & 1/4 & \infty & 1/4 & 1/4 & 0 \end{bmatrix} ,$$

where ∞ could be replaced by a sufficiently large value. Since D is an (6×6)-matrix, we may identify clusterings with partitions of $[6]$. Let μ represent the uniform distribution on $[6]$. It is easy to see that there are precisely two optimal 2-partitions,

$$\mathcal{C} = \{\{1,2,3\},\{4,5,6\}\} \text{ and } \mathcal{C}' = \{\{1,2,6\},\{4,5,3\}\} ,$$

each-one leading to risk 1/6. The gradients of $R(\mu,\mathcal{C})$ and $R(\mu,\mathcal{C}')$ satisfy

$$\nabla R(\mu,\mathcal{C}) = \nabla R(\mu,\mathcal{C}') = (1/4, 1/4, 0, 1/4, 1/4, 0)^\top .$$

The Hessians satisfy

$$\nabla^2 R(\mu,\mathcal{C}) = \nabla^2 R(\mu,\mathcal{C}') = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

Furthermore,

$$(\nabla R(\mu,\mathcal{C}))_{1,2,3} = -2 \neq 0 = (\nabla R(\mu,\mathcal{C}'))_{1,2,3} .$$

Thus the decision function $f(w) = R_{\mathcal{C}}(w) - R_{\mathcal{C}'}(w)$ has a vanishing gradient *and* a vanishing Hessian at $w = \mu$, but (as implied by our general discussion) the third-order term in the Taylor-expansion of f around μ does not vanish. For example, $(\nabla f(\mu))_{1,2,3} = -2$.