

Elements of Statistical Learning

Chapter 4, Exercise 1

Berit Omli Øksnes

February 2026

Suggested solution

This solution is based on pages 11-12 and 15-19 in these lecture slides: <https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec4RayleighQuotient.pdf>.

Assume that $\mathbf{B}, \mathbf{W}$ are symmetric matrices and that $\mathbf{W}$ is positive definite. We show how to solve the generalized eigenvalue problem

$$\arg \max_{\mathbf{a}} \mathbf{a}^T \mathbf{B} \mathbf{a} \text{ subject to } \mathbf{a}^T \mathbf{W} \mathbf{a} = 1, \quad (1)$$

by using the method of Lagrange multipliers. Define the Lagrangian function $\mathcal{L}$ by

$$\mathcal{L}(\mathbf{a}, \lambda) = \mathbf{a}^T \mathbf{B} \mathbf{a} + \lambda(1 - \mathbf{a}^T \mathbf{W} \mathbf{a}).$$

Then, we know that the optimal solution $(\mathbf{a}, \lambda)$ must satisfy $\frac{\partial \mathcal{L}}{\partial \mathbf{a}} = 0$ and $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$. We get that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{a}} &= 0 \\ \implies 2\mathbf{B}\mathbf{a} - 2\lambda\mathbf{W}\mathbf{a} &= 0 \\ \implies \mathbf{B}\mathbf{a} &= \lambda\mathbf{W}\mathbf{a} \\ \implies \mathbf{W}^{-1}\mathbf{B}\mathbf{a} &= \lambda\mathbf{a}, \end{aligned}$$

and further that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda} &= 0 \\ \implies 1 - \mathbf{a}^T \mathbf{W} \mathbf{a} &= 0 \\ \implies \mathbf{a}^T \mathbf{W} \mathbf{a} &= 1. \end{aligned}$$

We recognize the equation

$$\mathbf{W}^{-1}\mathbf{B}\mathbf{a} = \lambda\mathbf{a},$$

as a standard eigenvalue problem. That is, $\mathbf{a}$ must be an eigenvector of $\mathbf{W}^{-1}\mathbf{B}$ with corresponding eigenvalue λ . Further, if we multiply the above expression with $\mathbf{a}^T \mathbf{W}$ from the left, we find that

$$\begin{aligned} \mathbf{a}^T \mathbf{W} \mathbf{W}^{-1} \mathbf{B} \mathbf{a} &= \lambda \mathbf{a}^T \mathbf{W} \mathbf{a} \\ \implies \mathbf{a}^T \mathbf{B} \mathbf{a} &= \lambda, \end{aligned}$$

showing that for a given eigen pair $(\mathbf{a}, \lambda)$ such that $\mathbf{a}^T \mathbf{W} \mathbf{a} = 1$, it holds that $\mathbf{a}^T \mathbf{B} \mathbf{a} = \lambda$. Clearly, this expression is maximized when λ is the largest eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$. Hence, the solution to (1) is the eigenvector corresponding to the largest eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$.