Homework-4

1. Consider multivariate data $\mathbf{X}_{n\times p} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, where $\mathbf{x}_i \in \mathbb{R}^p, i = 1, 2, \ldots, n$, and denote the centered observations as $\mathbf{y}_i = \mathbf{x}_i - \bar{\mathbf{x}}$. Show that the criterion for the first sample principal component direction $\mathbf{a} = (a_1, a_2, \ldots, a_p)^T$, maximization of the sample variance of $\mathbf{a}^T\mathbf{y}_i$, is equivalent to minimization of the residual sum of squares, where the ith residual is defined as

$$\left(\mathbf{y}_i - \frac{\mathbf{a}^T\mathbf{y}_i}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\right)^T \left(\mathbf{y}_i - \frac{\mathbf{a}^T\mathbf{y}_i}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\right)$$

2. 5.22
3. 6.17
4. 6.22
5. 6.41
6. 7.21
7. 7.24
8. 7.27
9. **Reading assignments:** Chapters 6 to 8 of the textbook.