

MS-E2112 Multivariate Statistical Analysis – 2019

Exam

Answer to all the questions.

You are allowed to have pens and pencils, an eraser and a ruler, a calculator and one size A4 note (handwritten, text on one side only, name on the top right corner).

1. True or False (6 p.)

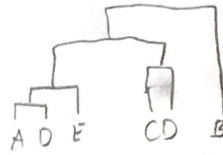
Determine whether the statement is true or false. (Every correct answer +1 p., every wrong answer -1 p., no answer 0 p.)

- (a) All affine equivariant scatter estimates do estimate the same population quantity.
- (b) In MCA, rare modalities have negligible/small effect on the analysis.
- (c) Canonical correlation analysis focuses on relationships between groups of variables.
- (d) Discriminant analysis is a method for splitting a set of individuals into unknown homogenous groups.
- (e) Fisher's linear discriminant analysis is based on maximizing the ratio of between groups dispersions to within group dispersions.
- (f) According to Zuo and Serfling, depth functions should be invariant under affine transformations.

Statement	a	b	c	d	e	f
True						
False						

Min:

Max: A — B



2. Clustering (6 p.)

Consider the following bivariate sample:

$$A = (2, 1), B = (-3, 0), C = (1, -2), D = (1, 1), E = (0, 2).$$

- Draw a scatter plot of the data. (1 p.)
- Perform agglomerative hierarchical clustering on the data. Use Euclidian distance as the distance measure and in clustering, measure the distance between the groups by applying maximum distance. Draw a corresponding classification tree. If you choose the number of the final clusters to be two, what are the two clusters? (5 p.)

3. Principal Component Analysis (6 p.)

Let $x \in \mathbb{R}^{p \times 1}$ be a p -variate random vector with mean μ and covariance matrix Σ . Let

$$\Sigma = \Gamma \Lambda \Gamma^T,$$

where the column vectors of Γ are orthonormal eigenvectors of Σ , and Λ is a diagonal matrix having its diagonal elements in decreasing order. Let $y = \Gamma^T(x - \mu)$, and let $z = a^T x$, where $a \in \mathbb{R}^{p \times 1}$ and $a^T a = 1$. Show that the variance of the first element of the vector y is larger than or equal to the variance of z .

4. Robust Methods (6 p.)

Describe how to measure robustness (locally and globally) and give examples of robust and non-robust measures of location and scatter.

BONUS QUESTION (2 p.):

Consider the following bivariate sample:

$$\{(4.5, 1.5), (-1.5, -2.5), (2.5, -1.5), (1.0, 1.0), (-0.5, 1.5), (0.0, 4.5)\}.$$

What is the half-space depth of the data point $(1.0, 1.0)$?