

R Course to VTT Technical Research Center of Finland

Exercise 2

Oguzhan GENCOGLU

Tampere University of Technology, VTT

May, 2016

Abstract

This document contains the tasks for Exercise 2. You can use the *Exercise_2_template.R* as a solution template. For questions, please contact `oguzhan.gencoglu@tut.fi`.

1 Task 1 - Lattice

Task 1 & 2.1 will use the same dataset, i.e., Breast Tissue dataset^{1 2}. Dataset with electrical impedance measurements of freshly excised tissue samples from the breast containing 106 observations of 11 features (attributes). Further information about the dataset can be found at <http://archive.ics.uci.edu/ml/datasets/Breast+Tissue#>

Try to replicate Figure 1 using the *lattice* plotting library. Plot *PA500* vs. *P* conditioned on *DA* and logarithm of *Area*.

Required functions:

- `read.xlsx`: Reads excel data. Use the *Data* sheet for the exercise.
- `equal.count`: Creates shingles objects. Create 2 shingles objects, i.e., one for *DA* other for logarithm of *Area*. Set the *number* argument to 3 for both.

¹Jossinet J (1996) Variability of impedivity in normal and pathological breast tissue. Med. & Biol. Eng. & Comput, 34: 346-350

²Silva JE, Marques de Sá JP, Jossinet J (2000) Classification of Breast Tissue by Electrical Impedance Spectroscopy. Med & Bio Eng & Computing, 38:26-30.

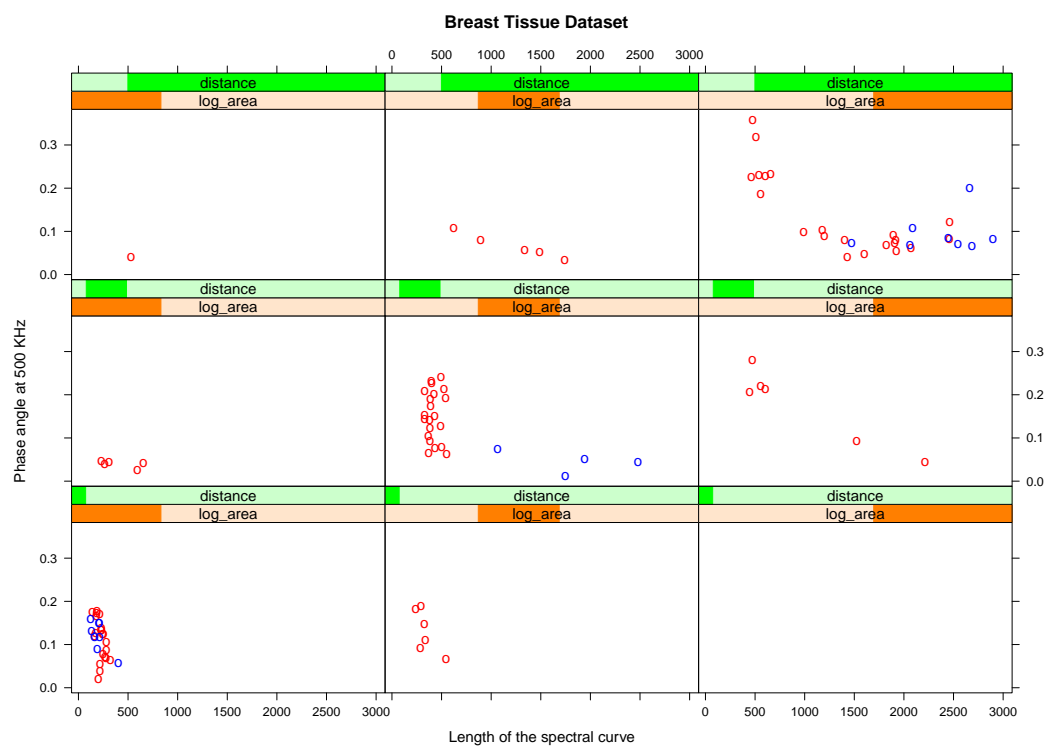


Figure 1: Task 1 - lattice plots.

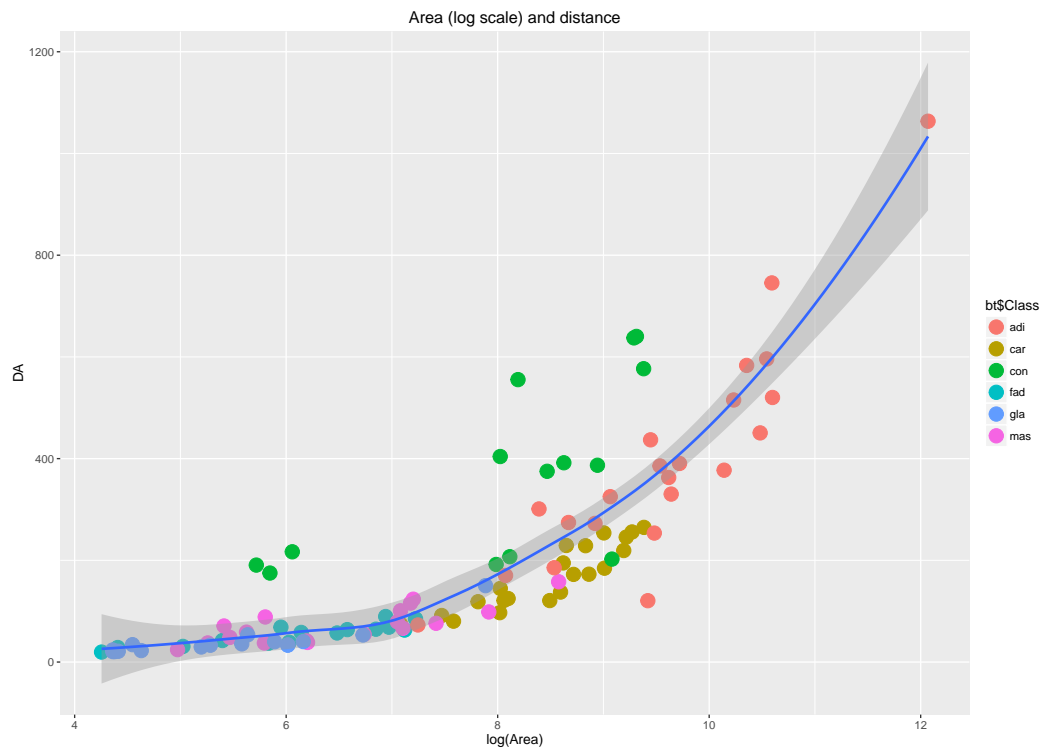


Figure 2: Task 2.1 - ggplot2.

- `xyplot`: Plots. Plot PA500 vs. P conditioned on logarithm of the area and distance. Use `col` argument to color the points with different *Class* categories.

2 Task 2 - ggplot2

2.1 Breast Tissue

Try to replicate Figure 2 using the *ggplot2* plotting library. Plot *DA* vs. logarithm of *Area*. Color the data points with the *Class* factor variable. Fit a curve as well.

Required functions:

- `ggplot`: Creates a canvas.
- `geom_point`: Adds a scatter plot.
- `geom_smooth`: Adds a fitted curve.

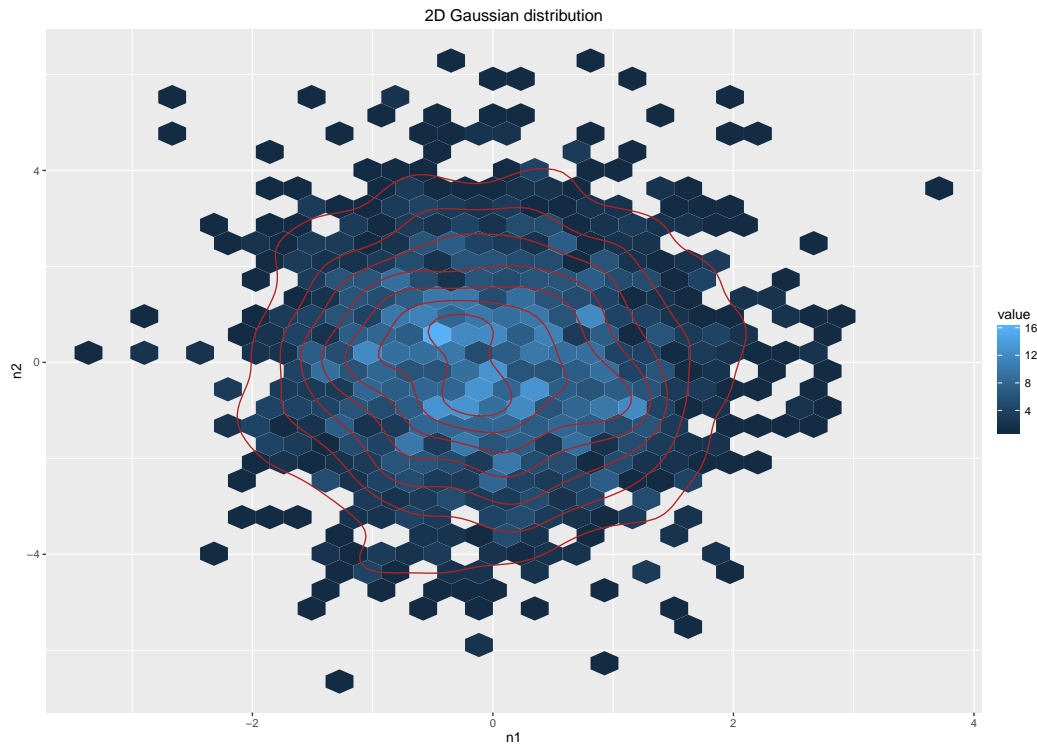


Figure 3: Task 2.2 - ggplot2.

- `ggtitle`: Adds title.

2.2 2D Gaussian Distribution

Try to replicate Figure 3 using the *ggplot2* plotting library. Create a 1500 row dataframe of 2 columns. First column consists of random samples from a zero mean, unity standard deviation Gaussian (normal) distribution. Second column is the same except that the samples would be from a distribution of standard deviation of 2.

Required functions:

- `data.frame`: Creates a dataframe.
- `rnorm`: Samples from a normal distribution.
- `ggplot`: Creates a canvas.
- `geom_hex`: Adds a hexagonal density plot.

- `geom_density2d`: Adds density curves.
- `ggtitle`: Adds title.