TDA231 - Algorithms for Machine Learning & Inference

Chalmers University of Technology

LP3– 2017-01-23

By:
Karabo Ikaneng, 891024-8239, kara.ikaneng@gmail.com
Elias Svensson, 920406-3052, elias.svensson.1992@gmail.com
Jacob Holmen, 940411-0372,jacobho@student.chalmers.se
Cong-Yi Xie, 911222-C552, goodtony314@gmail.com

Goal
Introduction to Probability, Matlab Primer.

1 Theoretical problems

1.1 Bayes Rule, 5 points

Information:

$$P(Test^+|Disease^+) = 0.99$$
$$P(Test^-|Disease^-) = 0.99$$
$$P(Disease^+) = \frac{1}{10000}$$

Quesiton:

$$P(Disease^+|Test^+) =?$$

Formulas:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{1}$$
$$P(A^+) = 1 - P(A^-) \tag{2}$$
$$P(A \cap B) = P(A|B) * P(B) \tag{3}$$

Solution:

$$P(Test^+|Disease^-) = 1 - P(Test^+|Disease^+) = 1 - 0.99 = 0.01$$
$$P(Test^-|Disease^+) = 1 - P(Test^-|Disease^-) = 1 - 0.99 = 0.01$$
$$P(Disease^-) = 1 - P(Disease^+) = 1 - \frac{1}{10000} = \frac{9999}{10000}$$
$$P(Disease^-) = 1 - P(Disease^+) = 1 - \frac{1}{10000} = \frac{9999}{10000}$$
$$P(Test^+) = P(Test^+|Disease^+) * P(Disease^+) + P(Test^+|Disease^-) * P(Disease^-) =$$
$$0.99 * \frac{1}{10000} + 0.01 * \frac{9999}{10000} \approx 0.010$$
$$P(Disease^+|Test^+) = \frac{P(Test^+|Disease^+) * P(Disease^+)}{P(Test^+)} =$$
$$\frac{P(Test^+|Disease^+) * P(Disease^+)}{P(Test^+)} = \frac{0.99 * \frac{1}{10000}}{0.010} \approx 0,0098$$

1.2 [Problem 1.2 [Correlation and Independence, 5 points]]

Information:

$$X \sim Uni([-1, 1])$$

$$Y := X^2$$

Question:

Show that $cov(X, Y) = 0$

Formulas:

$$cov(X, Y) = E[XY] - E[X]E[Y] \tag{1}$$

$$E[X] = \int_{-\infty}^{\infty} \mathrm{x} f(x)\, \mathrm{d}x \tag{2}$$

Solution:

$$E[X] = \int_{-1}^{1} \mathrm{x} \frac{1}{1 - (-1)}\, \mathrm{d}x = 0$$

$$E[XY] = \int_{-1}^{1} \mathrm{x}^3 \frac{1}{1 - (-1)}\, \mathrm{d}x = 0$$

$$cov(X, Y) = cov(X, X^2) = E[X * X^2] - E[X]E[X^2] = 0 - 0 * E[X^2] = 0$$

2 Practical problems

2.1 [Plotting normal distributed points, 5 points

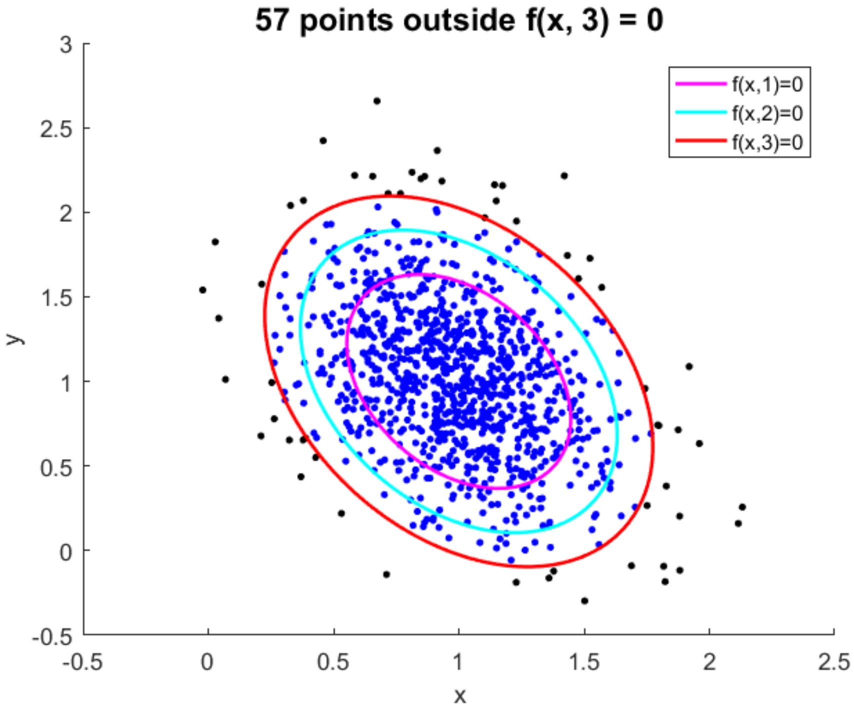

Figure 1: Level sets, scatter plot

The plot in Figure. 1 shows level sets $f(x, r) = 0$ for $r = 1, 2, 3$, title showing how many points lie outside $f(x, 3) = 0$, and a scatter plot of randomly generated points with points lying outside $f(x, 3) = 0$ showing in black while points inside shown in blue.

2.2 [Covariance and correlation, 5 points]

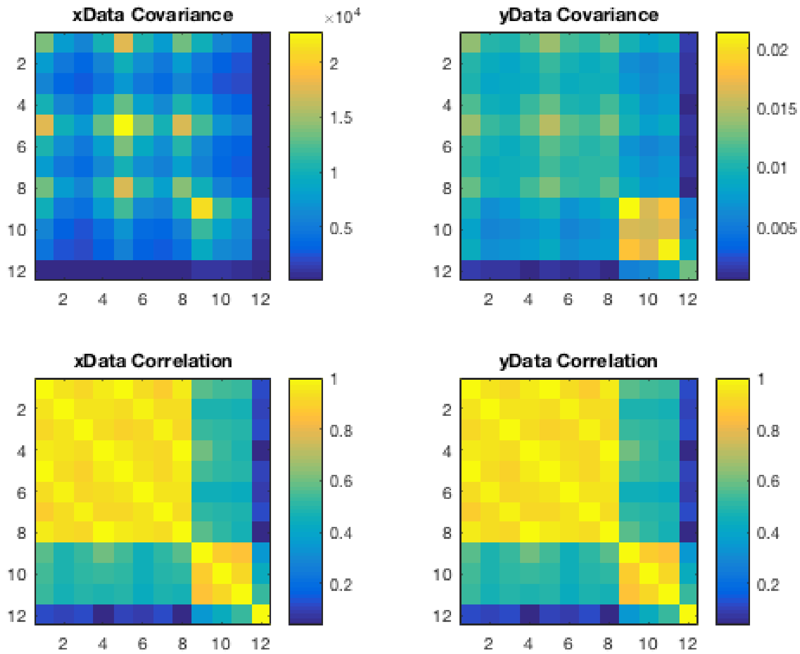

Figure 2: Colormap of correlation and covariance of X and Y

The plot in Figure. 2 shows colormaps of the covariance and correlation matrix for X, and Y which is scaled between [0,1]. One can observe that the correlation matrices of X and Y have the equivalent colormaps, but the covariance matrices of X and Y are different. Correlation does not change with scaling, and therefore remeain the same. However, the covariance of X is very large at points, but once the values are normalized to Y, naturally the covariance between the values become more similar to eachother, which is why the yData covariance plot has more similar colors than xData Covariance.

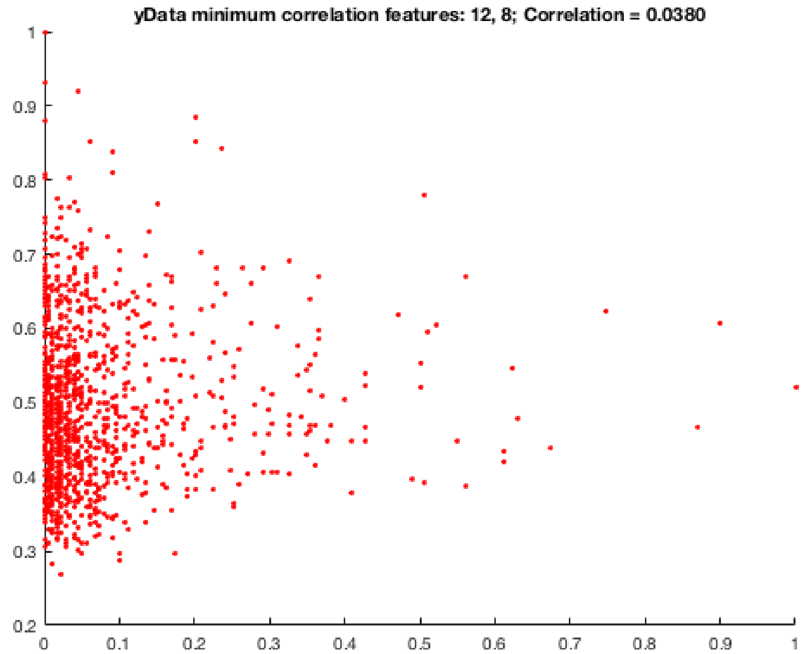

Figure 3: Features in Y having minimum correlation

The plot in Figure. 3 shows the pair of features in Y having minimum correlation, the features having indices 12 and 8, the correlation being 0.0380. Naturally, what this means is that among the 12 features in Y, features 12 and 8 correlate the least.