Change Detection with Diabetes using Graphical Lasso

Yeunkyung Kim and Yeboon Yun
Graduate School of Science and Engineering, Kansai University, Japan
e-mail: k545649@kansai-u.ac.jp, yeboon@kansai-u.ac.jp

Introduction

According to the demographic statistics of Japan for 2015, approximately 60% of total deaths are caused by the lifestyle-related diseases (Fig. 1).
In recent years, 'Data Health' has been increasingly conducted in Japan.
Using ***graphical lasso***, we try to analyze ***diagnosis data on diabetes***, and to ***detect an anomaly among factors*** with diabetes.

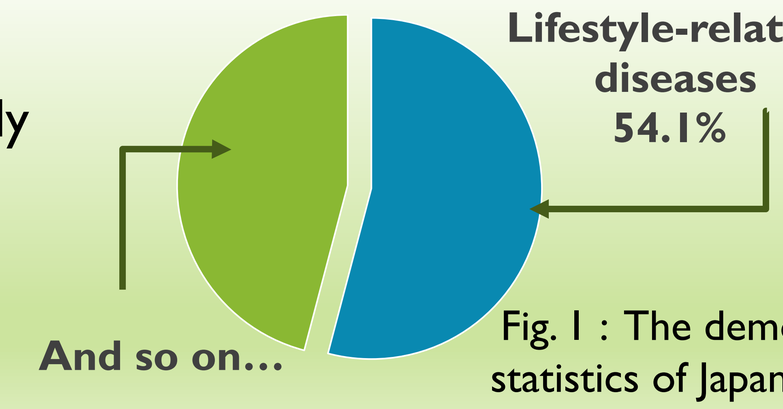


Fig. 1 : The demographic statistics of Japan for 2015

Sparse Learning Analysis by Graphical Lasso

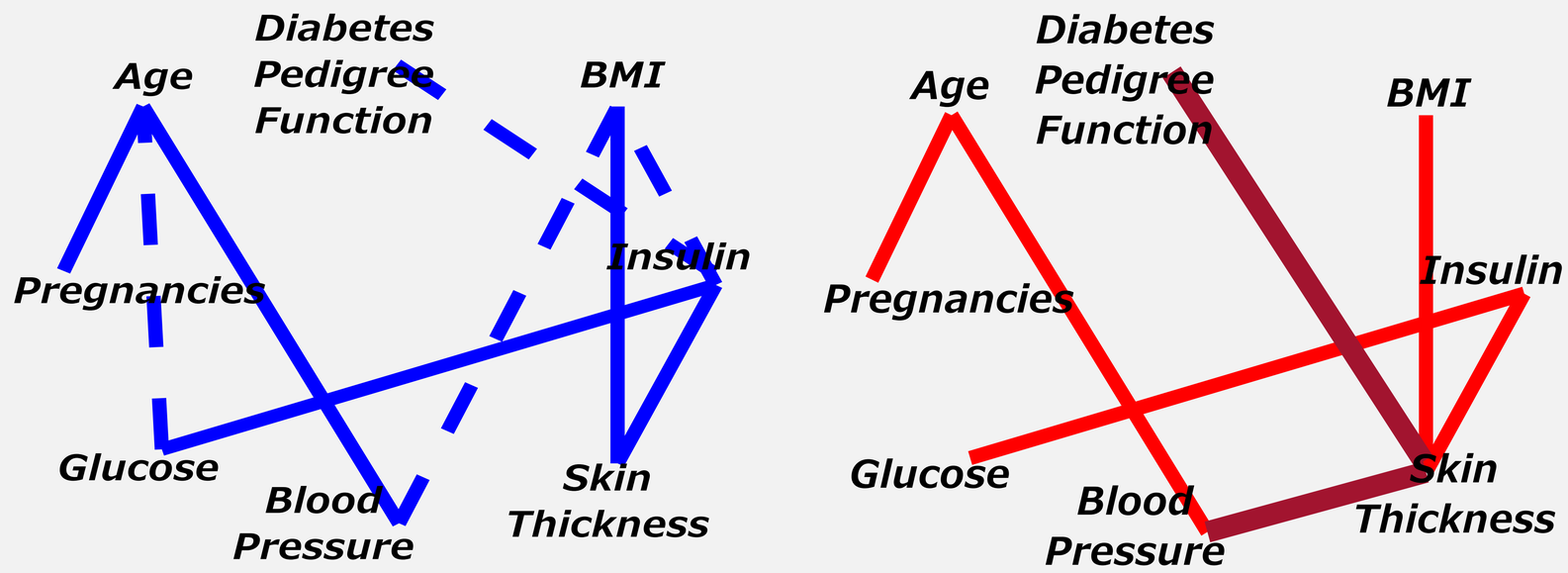
to estimate a precision matrix Λ of an inverse co-variance matrix as follows:

$$\Lambda^* = \arg\max_{\Lambda} \left(\ln\det\Lambda - \mathrm{tr}(S\Lambda) - \rho||\Lambda||_1\right)$$, where S : co-variance matrix for samples

$\rho > 0$: a regularization parameter

⇒ *Grasp of the features for two data sets based on a precision matrix Λ regraded as an adjacency matrix*

$$\Lambda_h{}^* = \begin{pmatrix} * & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & * & 0 & 0 & * & 0 & 0 & * \\ 0 & 0 & * & 0 & 0 & * & 0 & * \\ 0 & 0 & 0 & * & * & * & 0 & 0 \\ 0 & * & 0 & * & * & * & * & 0 \\ 0 & 0 & * & * & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & * & 0 \\ * & * & * & 0 & 0 & 0 & 0 & * \end{pmatrix} \quad \Lambda_p{}^* = \begin{pmatrix} * & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & * & 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & * & * & 0 & 0 & 0 & * \\ 0 & 0 & * & * & * & * & * & 0 \\ 0 & * & 0 & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & * & 0 \\ * & 0 & * & 0 & 0 & 0 & 0 & * \end{pmatrix}$$



(a) healthy individuals (b) patients with diabetes

Fig. 2 : Structure learning using graphical lasso

Table. 1 : Change degree

factor	degree
Pregnancies	0.039
Glucose	0.023
Blood Pressure	0.023
Skin Thickness	0.015
Insulin	0.014
BMI	**0.063**
Diabetes Pedigree Function	0.011
Age	0.037

based on the difference of the structures

Comparison with Support Vector Machines

to investigate the importance of each factor using SVM.

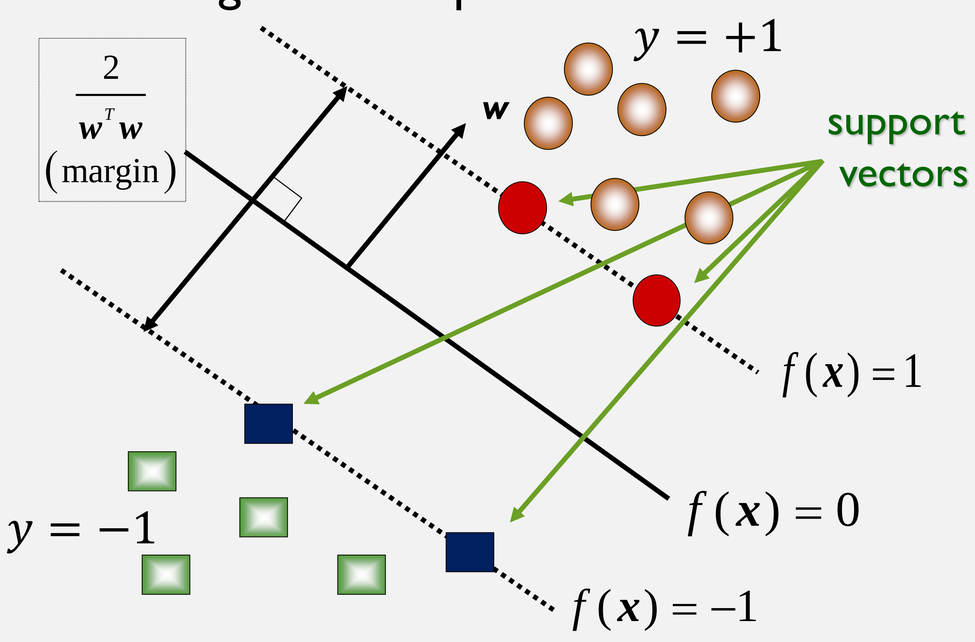
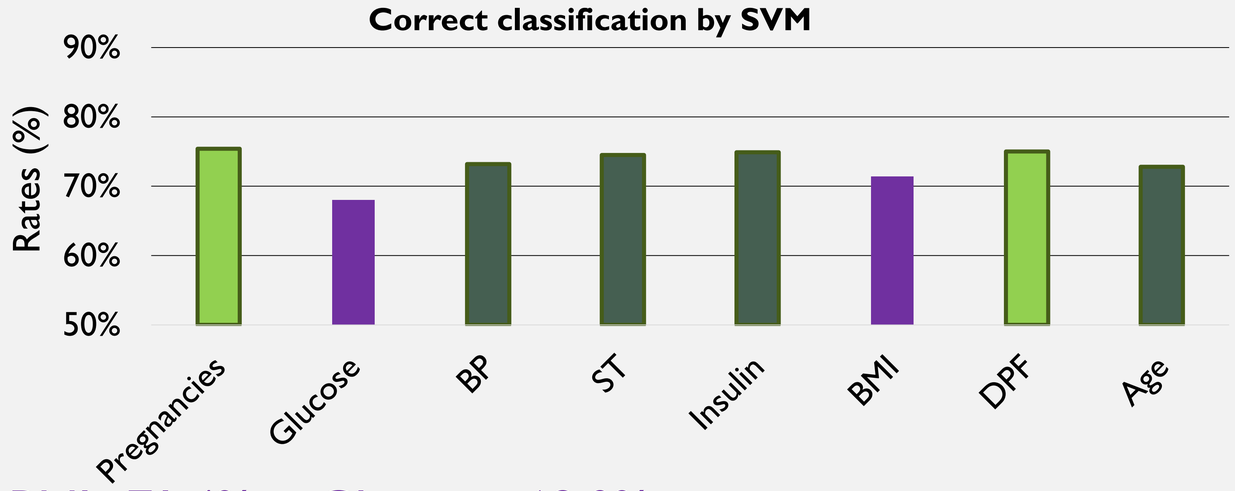


Fig. 3 : Overview of SVM



- BMI : 71.4% Glucose : 68.0%
  ⇒ important factor for diabetes diagnosis
- Pregnancies : 75.4% Diabetes Pedigree Function : 75.0%
  ⇒ not so important for diabetes diagnosis

* All : 74% (420 / 500 for healthy, 148 / 268 for patients)

Concluding Remarks

- The degree of change for BMI is the highest, which yields that BMI is important for diabetes diagnosis.
- The obtained results through this research will be expected to contribute for increasing efficiency in disease prevention and health promotion.