研究室名	小林靖之研究室　学会発表

【発表者について】アンダーラインは本学教員、研究員および技術職員、○は発表者、※は大学院生、卒研生または卒業生

学会名	The Conference of Data Science, Statistics & Visualisation (DSSV 2019) August 13-15, 2019 in Kyoto, Japan.
演題名	New Sparse Modeling of Sample Mahalanobis Distance
発表者	Yasuyuki Kobayashi
内容	Sparse modeling, such as Least Absolute Selection and Shrinkage Operator (LASSO) for regression has gained interest in variable selection to extract the essential data variables and prevent over-learning problems. Therefore, sparse modeling has also been applied to study the anomaly distance (AD). Thus far, only a sample covariance matrix S of learning samples x has been made sparse, for example, by applying graphical LASSO. However, the AD, such as the sample Mahalanobis distance (MD), of test sample y was not made sparse. Hence, this study was focused on making the AD of test sample y sparse. In principle, ordinal sample MD D^2 is given by D^2=(y-x ̄)^' S^(-1) (y-x ̄)=z'z, where x ̄ is the mean of the learning samples, and z is the studentized score vector (SSV) of y, i.e., z is the solution of linear equation y-x ̄=S^(1/2) z. I propose a new kind of sparse MD, D ̂^2, given by D ̂^2=z ̂^' z ̂, where z ̂ is the sparse solution of the equation obtained by applying the coordinate-decent method to solve LASSO. This sparse MD cancels the unstable effect of numerical error on the sample MD as follows. When learning samples x follow the p-variate normal distribution with population eigenvalues λ such that one λ_0=0 and the other λ>0 at the Monte Carlo simulation, sample eigenvalue l_0 of S corresponding to λ_0 becomes slightly positive under the influence of the numerical error, and D^2 becomes unstable owing to l_0. Subsequently, distributions of the element corresponding to l_0 of the SSV of test sample y were simulated as a(y), b(y), and c(y) for the ordinal, ridge, and sparse MDs, respectively. Here, $a(y)=((y-x ̄)·v_0)⁄√(l_0), b(y)=((y-x ̄)·v_0)⁄(√(l_0)+ρ), and c(y)=z ̂_((0)), where both x and y follow the same normal distribution with dimensionality p=7, v_0 is the sample eigenvector corresponding to l_0, regularizing constant ρ= 〖10〗^(-30)≅ l_0, and z ̂_((0)) is the element corresponding to l_0 of z ̂. The result shows that the sparse MD is useful for numerical computing. a(y) has a broad distribution because of a numerical error, and b(y) has a narrower distribution around y=0 than a(y). However, c(y) degenerates at y=0 correctly as λ_0=0, i.e., the effect of the numerical error is removed. However, for each of the other elements of the SSV of y, all the three distances show the same distribution.