이학석사 학위논문

Machine learning feature selection of financial data

(기계학습의 변수선별법을 통한 금융데이터 분석)

2015년 2월

서울대학교 대학원

수리과학부

김 시 연

Machine learning feature selection of financial data

(기계학습의 변수선별법을 통한 금융데이터 분석)

지도교수 최 형 인

이 논문을 이학석사 학위논문으로 제출함

2014년 12월

서울대학교 대학원

수리과학부

김 시 연

김 시 연의 이학석사 학위논문을 인준함

2014년 12월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

Machine learning feature selection of financial data

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Master of Science
to the faculty of the Graduate School of
Seoul National University

by

Shi Yeon Kim

Dissertation Director : Professor Hyeong-In Choi

Department of Mathematical Sciences
Seoul National University

February 2015

Abstract

Global financial crisis has been occurred frequently in these days also nations and companies pay attention to predict bankruptcy. In this thesis, we discuss feature selection method to extract feature group that causes main factors of making bankruptcy. We describe stepwise method and principal component analysis method briefly and compare it to construct prediction model. In addition, we try to analyze their performance and statistical measurement which method is the most efficient to raw data. We deal with data set of experiments which consist of 515 companies' financial statement in 1997 to build the model by using support vector machine.

Key words: feature selection, stepwise, principal component analysis, bankruptcy prediction
Student Number: 2012-20246

Contents

Chapter 1

Introduction

Bankruptcy prediction has been one of the most challenging task in many financial companies. Especially, it is one of main issue and topic in helping business loan in bank or making investment to firms. [1]
In 1997, I.M.F declared that South Korea had trouble in financial crisis, many Korean firms affected by the crisis in 1997. Also in 2010, Greek government-debt crisis is part of the ongoing Eurozone crisis, consecutively Italy, Portugal, and Spain are triggered bankruptcy of southern E.U. It has critically effected on financial problem to run firms moreover nations may be in danger of economic crisis. Incorrect decision making in financial information may the company troubled in distress to owner, shareholders, clients even government, etc. Also it spends time and social cost.[1] However, this is very complex and needs lots of information such as every prediction period. Accordingly, we have to choose more highly informative main factors of of bankruptcy to get more accuracy outcomes in financial decision making.
Recently, machine learning method is popular tools in large area to recognize pattern and time series forecasting due to its remarkable characteristics such as good generation performance. In developing a successful forecaster, feature selection is one of important step to improve performance better. [4]
In this paper, using the financial data of companies in 1994-1996, we do experiment about feature selection method to choose the best features and pattern for making prediction model more correctly. The contents of this paper are organized as follows.

In Chapter 2, we briefly describe the main feature selection method; step-

wise method and principle component analysis. We present how to reduce and extract features theoretically.

In Chapter 3, we provide the data and experiment procedures. We make prediction model to analyze in terms of preprocessing the feature selection method.

In Chapter 4, we will discuss the empirical result that we get. Classifying and comparing the group of stepwise and principal component, we analysis the main result for experiment.

In Chapter 5, we bring to the conclusion of our experiment in this paper.

Chapter 2

Feature Selection Method

In this section, we introduce representative feature selection techniques. Feature selection has been actively used in machine learning and statistic technique, such as data mining fields. It is important method to choose a group of set which is relevant features for constructing prediction models. There are many benefits of feature selection, which is improving model interpretability, defying the curse dimensionality to improve prediction model performance, facilitating data visualization, and reduce training times and cost. [1] We represent two feature selection methods; stepwise method and principal component analysis.

2.1 Stepwise method

Stepwise method is commonly used in feature selection method. This methods add and removes variable at each step. Before starting stepwise, we need to make the standard mutual information called entropy. That is mutual information.[1] Using probabilities models, we can match mutual information to statistical probabilities in rough.To select variables, stepwise must have t-test probability value, (in short, p- values) and p-value of an F-statistic is computed to set models. It needs to settle the marginal value of adding and removing probabilities we need, that is significant level. It's the combination of forward method and backward method. Before we describe stepwise

forward and backward method, let mutual information $Z_1, Z_2, ..., Z_d$, which related to output y.

(1) Stepwise forward method

Stepwise forward starts no features. It adds the features successively which match on mutual information (significant level) until enough variables are selected.

- Choose Z_i that satisfies

  $$I(Z_i, Y) = max_j I(Z_j, Y)$$

  which is the highest informative feature and numbering it as Z_{i_1}.

- Also choose Z_{i_2} which is the next highest feature. Repeat the steps to choose the features when ....

- Assume $Z_{i_1}, ...Z_{i_{l-1}}$ are selected. Add that satisfies

  $$I(Z_{i_l}, Y|Z_{i_1}, ..., Z_{i_{l-1}}) = max_j I(Z_j, Y|Z_{i_1}, ..., Z_{i_{l-1}})$$

  (Note : Since $I(X, Y|X) = 0$, there is no danger of choosing Z_{i_l} again among already chosen $Z_{i_1}, ..., Z_{i_{l-1}}$)

Stopping rules : Stop when there is no significant difference.

(2) Stepwise backward method

Stepwise backward starts all features eliminate successively which is highly uncorrelated on mutual information (significant level) until enough variables are eliminated.

- Eliminate irrelevent informative features Z_i that satisfies

  $$I(Z_i, Y|Z_k, k \neq \ i) = max_j I(Z_j, Y|Z_k, k \neq \ i)$$

  .

- Repeat the step until $Z_{i_1}, ..., Z_{i_{l-1}}$ are eliminated. Choose Z_{i_1} to be eliminated this time by

$$I(Z_{i_{il}}, Y|Z_k, k \neq \ i_1, \ldots, i_{l-1}, i_l) = min_j I(Z_j, Y|Z_k, k \neq \ i_1 \ldots, i_{l-1}, j)$$

- **Stopping Rules** : Stop when there is no significant difference

In short, we repeat the steps to eliminate the largest $p - value$ in small partial F-test when there is no significant difference between significant level. Also forward method repeat the steps adding variables become same as significant level. However, both method can't consider covariance between added or removed values. Hence that's the main drawback in this method.

For this matter, we use stepwise method which combine the advantage of two methods and that's the stepwise.

2.2 Principal component analysis

In this section, we introduce principal component analysis which widely used for dimensionality reduction, data compression also feature extraction.

Principal component analysis is the ways to find principal component in data set. It can be defined as the orthogonal projection of the data onto lower dimensional linear space, known as the principal subspace so we need to find the best plane that square error is minimum. [2]

Before starting, set $\mathscr{D} = (\mathbf{x}_i, y_i)_{i=1}^n$ be the data, and defined notation as random variable that represent data j^{th} feature Z_j, output y. From the Z_j, we define $n \times d$ design matrix $\mathbf{X}$ such as

$$\mathbf{X} = \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nd} \end{pmatrix}$$

To find the lower dimensional plane, we need to subtract the mean from data. First, define center mean vector, $\mu \in \mathbb{R}$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x_i}$$

and, write μ_j in component $\mu_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x_{ij}}$ for $j = 1, ..., d$.

Let normalize data matrix, call it new design matrix $\widetilde{\mathbf{X}}$,

$$\widetilde{\mathbf{X}} = \begin{pmatrix} x_{11} - \mu_{11} & \dots & x_{1n} - \mu_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} - \mu_{n1} & \dots & x_{nd} - \mu_{nd} \end{pmatrix}$$

Figure 2.1 shows an example of calculating the mean of datas if two groups of feature sets exist.

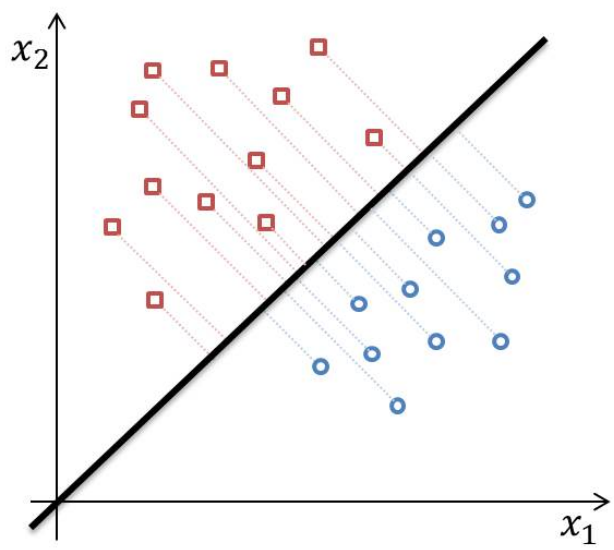

Figure 2.1

For representing data set which is most reserved, it is the best way to find the direction that data spread out the most. We need to find the maximum

vector that capture the shape of data most as figure 2.2.

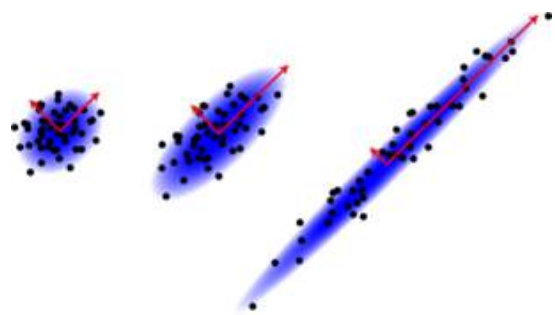

Figure 2.2

So we need to find unit vector $v \in \mathbb{R}$,
$F(v) = \sum_{l=1}^{n} |proj_v(\mathbf{x}_l - \mu)|^2 = \sum_{l=1}^{n} |v^T(\mathbf{x}_l - \mu)|^2$ is maximum.

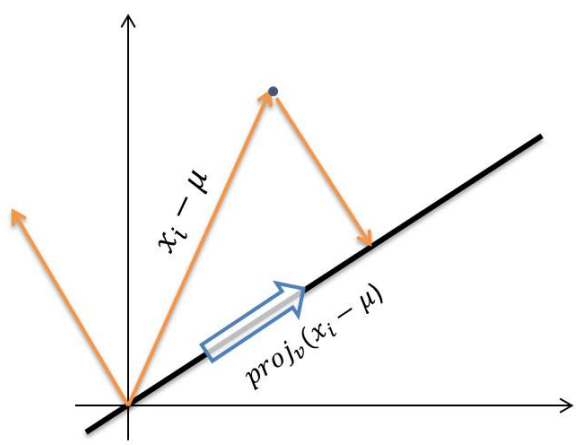

Figure 2.3

Calculating $F(v)$

$$F(v) = v^T\{(\mathbf{x}_l - \mu)(\mathbf{x}_l - \mu)^T\}v$$

Define symmetric covariance matrix C, $C = \sum_{l=1}^{n}(\mathbf{x}_l - \mu)(\mathbf{x}_l - \mu)^T$ which is $d \times d$.

Using Lagrange multiplier to find maximize $F(v)$,

$$\mathbf{L}(\mathbf{v}, \lambda) = \sum_{i=1}^{n} |v^T(\mathbf{x}_l - \mu)|^2 - \lambda(|v|^2 - 1) = \sum_{i,j=1}^{d} v_i v_j c_{ij} - \lambda\{\sum_{i=1}^{d} v_i^2 - 1\}$$

Then, we get $cv = \lambda v$. v must be an eigenvector of c. Using diagonalization theorem, we represent symmetric matrix C as $C = \lambda_1 v_1 v_1^T + \cdots + \lambda_d v_d v_d^T$ where $\{v_1 \ldots v_d\}$ an orthonormal basis of $\mathbb{R}$ and $\lambda_1 ... \lambda_d$ are eigenvalues of C ordered in the descending order : $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_d \geq 0$. We can easily check $F(v)$ is the biggest when $v = v_1$. [2] Also this eigenvalues turn out that the eigenvector with the highest eigenvalues are the principal component of the data set. Because of representing the most effective way, we need to choose the largest value. More generally, we consider multi-dimensional version.

Let $F(w) = \sum_{i=1}^{n} |proj_w(x_l - \mu)|^2$ be the projection of the normalized data onto the subspaces w. If w is the subspace spanned by k the highest eigenvector, $F(w)$ is maximum among all k-dimensional linear space of $\mathbb{R}$.

This means k dimensional linear space represents the approximated data of k. By using singular value decomposition, we factorize the matrix to find the eigenvalue-eigenvector; k-dimensional subspace more easily.

Singular Value Decomposition

Singular value decomposition in $\mathbb{R}$, $m \times n$ matrix A can be factored into,

$$\mathbf{A} = \mathbf{UDV^T}$$

U : $m \times m$ orthogonal matrix

V : $n \times n$ orthogonal matrix

D : $m \times n$ diagonal matrix

Singular value decomposition is closely associated with eigenvalue-eigenvector factorization of positive definite matrix.[6] Diagonal elements of D means singular values of A which means square root eigenvectors of AA^T and A^TA. Column vector of U represents left singular vectors of A as eigenvector of AA^T and column vector of V represents right singular vector of A, that means eigenvector of A^TA.

Applying Singular Value Decomposition, we note that the $\widetilde{X} = UDV^T$, $n \times d$ matrix also represent C as

$$C = \widetilde{X}^T\widetilde{X} = VD^TDV^T = V\Lambda V^T \text{ where} \Lambda = D^TD \text{ is } d \times d. \; C\Lambda = V\Lambda.$$

Let $e_i = (0 \dots 1 \dots)^T$ be the d dimensional standard basis vector. Note that $Ve_j = v_j$. Thus $CVe_j = Cv_j$ $V\Lambda e_j = \lambda_j Ve_j = \lambda v_j$. [2] Next, we make new features which are chosen from original one. Let R as a matrix $n \times d$.

$$\mathbf{R} = \mathbf{UD} = \begin{pmatrix} r_{11} & \dots & r_{1d} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{nd} \end{pmatrix}$$

We have $\widetilde{X} = RV^T$. Let $\widetilde{e}_i$ be the n dimensional standard basis vector whose components are all zero except the ith which is 1. Then we easily check the ith row vector of $\widetilde{X}^T$ is exactly same as the column vector. [2]

$$R^T\widetilde{e}_i = \sum_{j=1}^{d}(R^T_{ji}e_j) = \sum_{j=1}^{d} r_{ij}e_j.$$

Thus

$$\widetilde{X}^T = VR^T\widetilde{e}_i = \sum_{j=1}^{d} r_{ij}Ve_j = \sum_{j=1}^{d} r_{ij}v_j$$

From feature d-dimensional subspace, we determine the feature number k which projected to W. This k-features become new feature as mentioned.

Once the number of eigenvector k that spread the figure most is chosen, we make the new feature data. In d-dimensional subspace, we define the degree of closeness, $Tr(C) = \lambda_1 + \cdots + \lambda_d$ and capture the linear space. For capturing the subspace W, select k eigenvectors from $Tr(C)$,degree of closeness which cut off at k so that $\lambda_1 + \cdots + \lambda_k$. If we capture 50% subspace of $Tr(C)$, choose the smallest k $\lambda_1 + \cdots + \lambda_k \geq 0.5 \cdot Tr(C)$.

Chapter 3

Data and experiment

3.1 Data

Data set includes bankruptcy and non-bankruptcy financial statements of Korean companies in 1994-1996. We collect 515 companies, based on balance sheet and profit and loss account. 133 companies had been bankrupted in 1997, and the rest of those had not. We establish features based on financial statements and can classify criteria such as stability, liquidity, asset quality etc.
As shown in Table 3.1, we construct 56 feature values to construct prediction model. [2] [10]

3.2 Experiment

In this experiment, we used feature selection tools, stepwise method and principal component analysis previously mentioned in chapter 2. By classifying the degree of closeness, we denote P.C.A as 3 classes; capturing the data 100% denoted $PCA1$, 99% denoted $PCA2$ and 90% as $PCA3$.
In stepwise method, we set significant level enter $\alpha_E = 0.1, \alpha_R = 0.1$. To analysis more carefully, we construct top 5 features which chosen from stepwise, named as Stepwise 5. We distinguish the bankrupted data output y_i as 0 and 1 ; 0 is 'non-default' companies, and 1 is 'default'.

The classification method in machine learning we used is support vector machine with supervised learning algorithm. We use 10% cross validation, for dividing data 90% as a training and 10% as testing set. The kernel function we use in the support vector machine is Gaussian kernel, which is $\phi(x_i)\phi(x_j) = exp(-\gamma\big||x_i - x_j|\big|^2)$.

For evaluating correctness of the performance, we use statiscal measurement based on the confusion matrix in Table 3.2.[9]

TP : $TruePositive$; the model predicts correctly when it's a positive. TN is $TrueNegative$; the model predicts it's negative. Also FP and FN is $FalsePositive$ and $FalseNegative$ which predict actual class wrongly.
The statistical measures are as following:

(1) Accuracy = (TN+TP)/(TN+FN+FP+TP)

(2) Recall = TP /(FN+TP)

(3) Specificity = TN /(FP +TN)

(4) Precision (Positive predictive value) = TP /(FP + TP)

(5) Negative prediction value = TN / (FN + TN)

(6) Type 1 error = 1-specificity

(7) Type 2 error = 1- Recall

Number	Feature Information
1	Total Asset in 1996
2	Total Liability in 1996
3	Debt to equity ratio in1996
4	Debt ratio in 1996
5	Current ratio in 1996
6	Times interest earned in 1996
7	Operating profit margin in 1996
8	Fixed assets to stockholder's equity and long-term liabilities in 1996
9	Quick ratio in 1996
10	Return on total assets in 1996
11	Return of equity in 1996
12	Net income before tax in 1996
13	Earning per share in 1996
14	Gross profit margin in 1996
15	Net profit in 1996
16	Net income before tax / sales in 1996
17	Cash flow from operating activities in 1996
18	Cash flow from operating profit in 1996
19	Relative difference of liabilities between 1994 and 1996
20	Relative difference of liabilities between 1995 and 1996
21	Relative difference of net profit between 1994 and 1996
22	Relative difference of net profit between 1995 and 1996
23	Relative difference of capital between 1994 and 1996
24	Relative difference of capital between 1995 and 1996
25	Relative difference of total assets between 1994 and 1996
26	Relative difference of total assets between 1995 and 1996
27	Relative difference of gross profit between 1994 and 1996
28	Relative difference of gross profit between1995 and 1996
29	Relative difference of inventory 1994 and 1996
30	Relative difference of inventory 1995 and 1996
31	Relative difference of cash and deposit between 1994 and 1996
32	Relative difference of cash and deposit between 1995 and 1996
33	Capital reserve ratio
34	Operating profit to net sales ratio
35	Rate of sales cost
36	Sales cost in 1995

37	Relative difference of current ratio between 1994 and 1996
38	Relative difference of current ratio between 1995 and 1996
39	Current ratio in1995
40	Relative difference of inventory turnover ratio between 1994 and 1996
41	Relative difference of inventory turnover ratio between 1996 and 1996
42	Relative difference of growth of net profit rate between 1994 and 1996
43	Relative difference of growth of net profit rate between 1995 and 1996
44	Relative difference of net interest margin between 1994 and 1996
45	Relative difference of net interest margin between 1995 and 1996
46	Relative difference of receivable turnover ratio between 1994 and 1996
47	Relative difference of receivable turnover ratio between 1995 and 1996
48	Quick ratio in 1995
49	Relative difference of tangible fixed assets turnover between 1994 and 1996
50	Relative difference of tangible fixed assets turnover between 1995 and 1996
51	Relative difference of working asset ratio between 1994 and 1996
52	Relative difference of working asset ratio between 1995 and 1996
53	Total borrowings to total asset in 1995
54	Total borrowings to total asset in 1996
55	Relative difference of tangible asset between 1994 and 1996
56	Relative difference of tangible asset between 1995 and 1996

Table 3.1: Feature

		Actual class	
		0	1
predicted	0	TN	FN
Class	1	FP	TP

Table 3.2: Confusion matrix

Chapter 4

Result

In this chapter, we discuss the result of this experiment. By comparing the performance of feature selection, we want to find the effect ion of feature selection to build prediction model between raw data and selected data.

4.1 Stepwise method

In stepwise method, it select relevant features directly to compare significant level. As you see Table 4.1, 11 features are progressively chosen by this method.

	Feature
x4	Debt ratio in 1996
x52	Relative difference of increasing tangible asset between 1994 and 1996
x25	Relative difference of total assets between 1995 and 1996
x10	Return on total asset in 1996
x44	Relative difference of net interest margin between 1994 and 1996
x46	Relative difference of receivable turnover ratio between 1994 and 1996
x41	Relative difference of growth rate of net income between 1995 and 1996
x12	Net income before tax in 1996
x53	Relative difference of increasing tangible asset between 1995 and 1996
x54	Relative difference of increasing tangible asset between 1994 and 1996
x36	Relative difference of inventory between 1995 and 1996

Table 4.1: Results

By selecting ranked 5 features using stepwise method, we analysis statical measurement more sophisticated, and calls it as stepwise 5. For examining the selected data, we make fake data set which are irrelevant to this stepwise method. We call it fake data.

As shown in Table 4.2, we compare the result which comes out from stepwise. Overall, stepwise 11 is the best performance. Average accuracy, recall and precision are getting higher when the number of features becomes larger. Also stepwise method selects 11 features and extract ratio is 78.57%.

	Fake data	Stepwise 5	Stepwise11
Accuracy	66.62%	71.4 %	76.42%
Recall	15.03%	25.75%	47.04%
Specificity	84.59%	87.30%	86.64%
Precision	25.37%	41.37%	55.10%
Negative predictive value	74.09%	77.15%	82.45%
Type 1 error	15.4%	12.69%	13.35%
Type 2 error	84.96%	74.24%	52.95%

Table 4.2: Results

4.2 Principal component analysis

In principal component analysis, we classify 3 class to compare the data. As shown in Table 4.3, the average accuracy of PCA are 69.87% and it doesn't have much difference between each other. Comparing the percentage of recall and precision, we find $PCA2$ is as twice as PCA. Even though Type 1 error becomes higher when we set lower closeness to original principal subspace , we can reduce Type2 error dramatically. Principal component analysis qualifies the data to improve most by checking the percentage of Type 1 and 2 errors.
The calculated the ratio of selected feature numbers are in Table 4.4. When closeness goes higher to principal component subspace, the number of selected

features become larger. PCA selects 56 eigenvectors, $PCA1$ is 42 and $PCA2$ is 23. We also check extraction rate and accuracy. $PCA2$ extract the data almost half of them, 23 features. Accuracy has not influence in principal component analysis.

	PCA	$PCA1$	$PCA2$
Accuracy	68.71%	70.12%	70.79%
Recall	12.67%	36.94%	44.89%
Specificity	88.21%	81.67%	79.80%
Precision	27.24%	41.30%	43.63%
Negative predictive value	74.36%	78.81%	80.63%
Type 1 error	11.78%	18.32%	20.19%
Type 2 error	87.32%	63.05%	55.10%

Table 4.3: Result :Principal component analysis

	PCA	$PCA1$	$PCA2$
Number of Selected variable	56	42	23
extraction rate	0%	25%	46.4%

Table 4.4: Result : Selected feature in Principal component analysis

4.3 Comparing feature selection

Table 4.5 shows the performance of using raw data, stepwise and principal component analysis. Regarding Table 4.5, stepwise is the first highest accuracy 76.42%, next is $PCA2$, 70.79%. Recall of stepwise and $PCA2$ are almost 4 times as much as raw data set. Likewise, in statistical measurement, feature selection improves performance of support machine to build the prediction model. Concerning the effectiveness of using feature selection method, we examined the extraction rate to compare the result. Stepwise is outperformed in extraction rate 75% and $PCA2$ is the 53.6%. Actually, it is insignificant

to comparing the feature extraction rate when it comes to effectiveness of feature selection method. However, it could make an effects on the other factors when construct the model such as time and expenses.

	Raw data	Stepwise	$PCA2$
Accuracy	66.40%	76.42%	70.79%
Recall	11.06%	47.04%	44.89%
Specificity	85.67%	86.64%	79.80%
Precision	20.92%	55.11%	43.63%
Negative predictive value	73.45%	82.45%	80.63%
Type 1 error	14.32%	13.35%	20.19%
Type 2 error	88.93%	52.95%	55.10%

Table 4.5: Result : Comparing Raw data to stepwise and P.C.A

	Stepwise	PCA	$PCA2$
Accuracy	76.42%	70.12%	70.79%
extraction rate	78.57%	25%	46.4%

Table 4.6: Result : Feature extraction rate

Following the Figure 4.2, type 1 and 2 errors are 14.32% and 88.93% in raw data, remarkably imbalanced than others. After trimming data, it shows percentage of error is significantly improved. Unlikely stepwise type 1 error slightly gets worse, but type 2 error gets better. It implies the preprocessing the data make the quality of data set better. The following figures from 4.1 to 4.3 are the comparing the results in graphs.

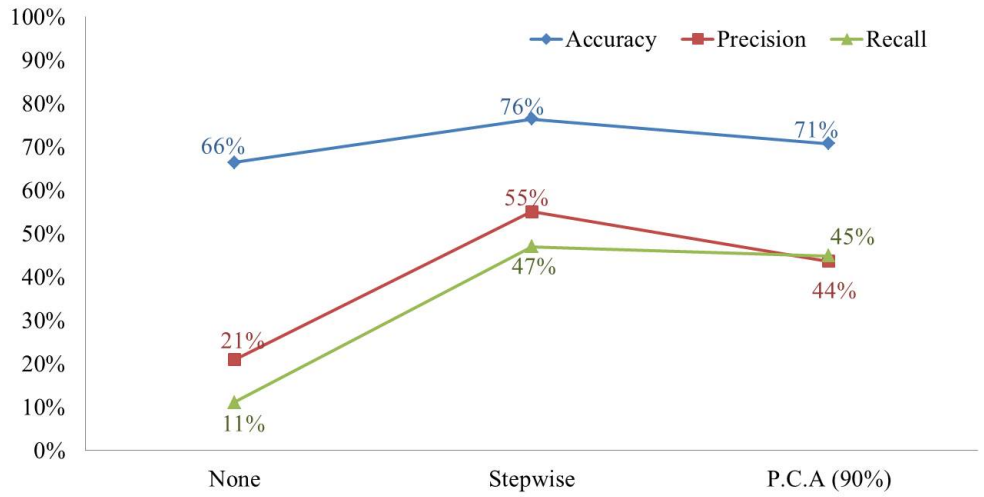

Figure 4.1: Result

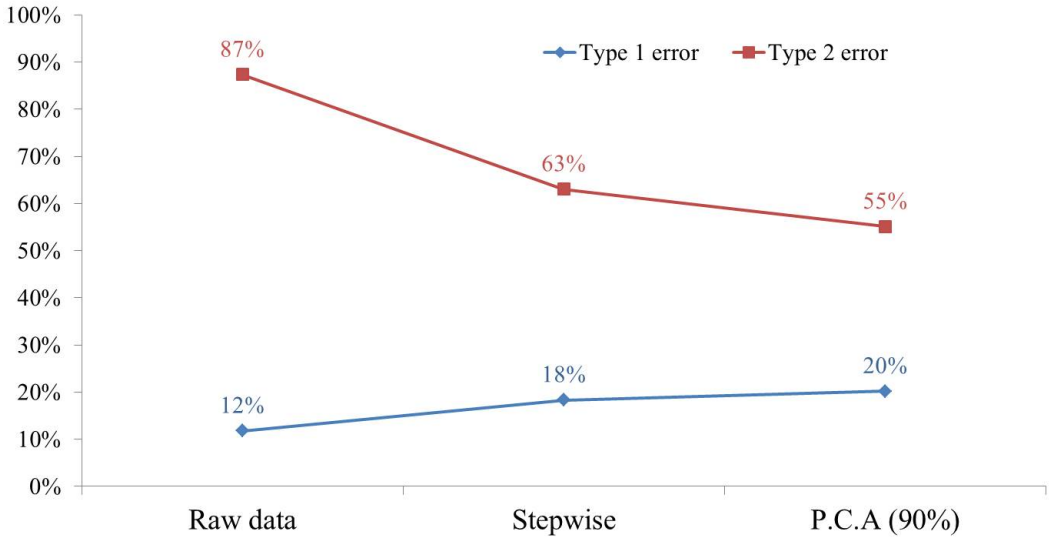

Figure 4.2: Result : Type 1 and 2 errors

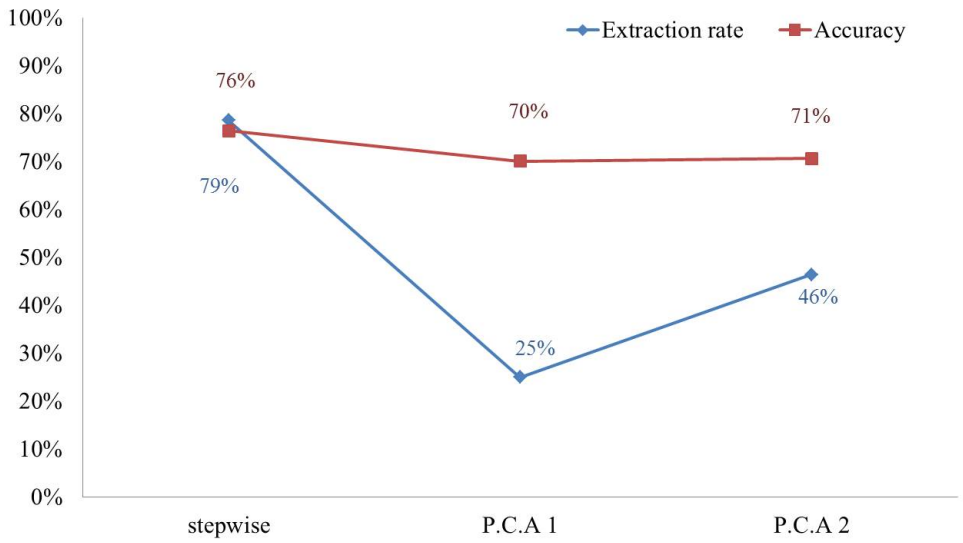

Figure 4.3: Result : Accuracy and Feature extraction rate

Chapter 5

Conclusion

As we applied feature selection method, we compared the performance of raw data and processed data, such as stepwise and principal component analysis. Regarding this experiment and result, we found that stepwise method makes more superior performance than principal component analysis overall. In statical measurement, accuracy and contraction rate. In this experiment stepwise method build prediction model more better. Comparing the average accuracy, principal component analysis seems that it doesn't have much difference between raw data. However, type 1 and 2 errors reduced remarkably, we get qualitative improvement after trimming principal component analysis. So it make imbalanced data to perform discriminate effectively. In the closeness to principal component subspace, we get lower contraction rate when setting higher closeness to subspace.
For this experiment we could say that feature selection, pre-processing step, is important when constructing prediction model, especially large amount of information in highly correlated variables.

In this paper, we only focus on 1994 to 1996 financial statement in Korean company. If we considered other data set, for instance more various nation, credit data or other period of financial data, we could get more stable and correct outcomes for this experiment. For future work, adding other related business field, such as exchange rates, stock price, oil price could get more selected feature which is less correlated.

Bibliography

[1] Chi-Fong Tsai, *Feature selection in bankruptcy prediction*, Knowledge-Based Systems 22, 2009.

[2] H. I. Choi, *Feature selection and Extraction*, lecture note, preprint, 2014.

[3] Chistopher M.Bishop, *Pattern recognition and Machine learning*, Spinger, 2009.

[4] L.J Cao et el, *A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine.*

[5] Frenando Mendes et el, *Feature Selection for Bankruptcy Prediction: A Multi Objective Optimization Approach.*

[6] Gilbert Strang, *Linear algebra and its application*, Thomson Brooks/Coles, Forth edition.

[7] Gang Wang et el, *An improved boosting based on feature selection for corporate bankruptcy prediction*, 2014.

[8] Edward I. Altman, *The journal of Finance*, September 1964.

[9] Ahran Lee, *Default Prediction by Using Machine Learning Methods*, Seoul National University, .

[10] Sang Bong Kim, Philip Ji, Kyung-june Jo, *부도예측모형을 이용한 기업 부실화의 원인분석*, 시장경제연구 40집 1호.

[11] Andrew Eg, *CS229 : Lecture Notes, Principal component analysis*,p Machine learning lecture notes.

국문초록

세계적 금융 위기로 인해 국가는 물론 기업의 재무 건전성에 대한 관심은 높아졌고 부도율 예측에 대한 관심이 급증되었다. 이 논문에서는 부도에 영향력 높은 변수를 선별하고 압축하는 변수 선별법에 대해 논한다. 스텝와이즈와 주성분 분석을 소개 및 비교하고 어느 방법이 더 효율적으로 변수를 압축하는지 로우 데이터와 비교분석하여 효율성을 알아본다. 1997년 515개의 기업 재무제표 및 손익계산서를 토대로 변수를 설정하여 서포트 벡터 머신을 이용해 모델을 구축한 후, 실험을 하였다.

주요어휘: 변수 선별법, 스텝와이즈, 주성분 분석, 부도율 예측
학번: 2012-20246