STATISTICS and its INTERACTIONS with OTHER DISCIPLINES (SIOD 2013)
June 4 - 7, 2013, Ho Chi Minh, Vietnam

A new method to determine separated hyper-plane for non-parametric sign test in multivariate data

Loc Nguyen, Post-Doctoral Researcher
Vietnam Institute of Mathematics
Email: ng_phloc@yahoo.com
Homepage: www.locnguyen.net

Abstract

Non-parametric testing is very necessary in case that the statistical sample does not conform normal distribution or we have no knowledge about sample distribution. Sign test is a popular and effective test for non-parametric model but it cannot be applied into multivariate data in which observations are vectors because the ordering and comparative operators are not defined in n-dimension vector space. So, this research proposes a new approach to perform sign test on multivariate sample by using a hyper-plane to separate multi-dimensional observations into two sides. Therefore, it is possible for the sign test to assign plus signs and minus signs to observations in each side. Moreover, this research introduces a new method to determine the separated hyper-plane. This method is a variant of support vector machine (SVM), thus, the optimized hyper-plane is the one that contains null hypothesis and splits observations as discriminatively as possible.

Keywords: separated hyper-plane, non-parametric sign test.

Table of contents

1. Introduction to non-parametric sign test

- Nonparametric testing is used in case of without knowledge about sample distribution; concretely, there is no assumption of normality.
- The nonparametric testing begins with the test on sample *median* in univariate data. If distribution is symmetric, median is identical to mean.
- Given the median μ_0 is the observation at which the left side data and the right side data are of equal accumulate probability: $P(D < \mu_0) = P(D > \mu_0) = 0.5$.

1. Introduction to non-parametric sign test

The sign test [1, pp. 656-660] is performed as below :

- Assigning plus signs to sample observations whose values are greater than and minus signs to ones whose values are less than.
- If the number of plus signs is nearly equal to the number of minus signs, then null hypothesis H_0 is true; otherwise H_0 is false. In other words, that the proportion of plus signs is significantly different from 0.5 cause to rejecting H_0 in flavor of H_1.

2. Multivariate non-parametric sign test

Basic idea

- Traditional sign test cannot be applied into multivariate data because null hypothesis and alternative hypothesis are vectors and it is not likely to compare hypothesis vector with observation vector with regard to assign plus signs.
- The basic idea is to determine *a hyper-plane and its normal vector which is used to calculate signed-rank number*.
- The optimized hyper-plane is the one that contains null hypothesis and splits observations as discriminatively as possible.

2. Multivariate non-parametric sign test

Multivariate non-parametric sign test includes three steps

1. Determining *separated hyper-plane* that contains null hypothesis μ_0 and splits observations as discriminatively as possible.
2. The normal vector W of separated hyper-plane is used to make sign of observations. Observations having the same direction to W are assigned plus sign and otherwise. The method to discover this hyper-plane is described particularly in section 3.
3. Null hypothesis μ_0 is rejected or accepted in flavor of alternative hypothesis based on the number of plus signs.

3. Determining separated hyper-plane

Given sample $\Omega = \{X_1, X_2, \ldots, X_m\}$ has m observations, the **separated hyper-plane** satisfies two following propositions:

1. It contains specified null hypothesis μ_0, which means:
$$1 - |W^T(X_i - \mu_0)| \leq 0$$
The first proposition is similar to the methodology of support vector machine (SVM) method [2].
2. It splits vector space Ω into two half-spaces as separately as possible, which means W is extreme point of following Lagrange function [3, p. 215]:

$$L(W, \lambda) = \frac{1}{2}W^TW + \sum_{i=1}^{m}\lambda_i(1 - |W^T(X_i - \mu_0)|)$$

3. Determining separated hyper-plane

- Similar to SVM method [2], by solving optimization problem, the Lagrange function becomes:

$$L(W,\lambda) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j \frac{W^T(X_i-\mu_0)}{|W^T(X_i-\mu_0)|}\frac{W^T(X_j-\mu_0)}{|W^T(X_j-\mu_0)|}(X_i-\mu_0)^T(X_j-\mu_0)+\sum_{i=1}^{m}\lambda_i - \sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j\left|(X_i-\mu_0)^T(X_j-\mu_0)\right|$$

- By some arithmetic transformations, we get lower bound $L_1(\lambda)$ and upper bound $L_1(\lambda)$ of $L(W, \lambda)$ so that $L_1(\lambda) \le L(W, \lambda) \le L_2(\lambda)$. We have:

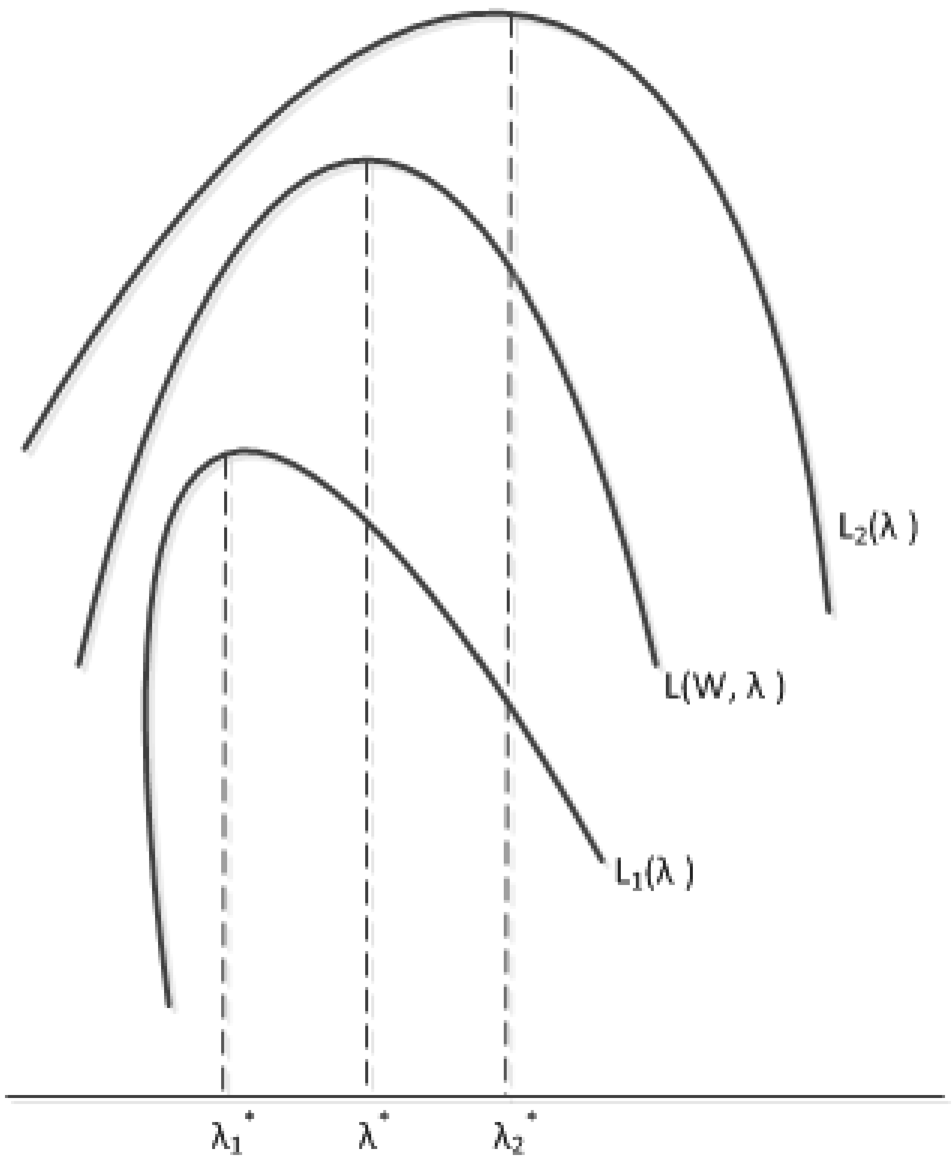
$$L_1(\lambda) = \lambda^T I - \frac{3}{2}\lambda^T S\lambda$$

$$L_2(\lambda) = \lambda^T I - \frac{1}{2}\lambda^T S\lambda$$

Where $I = (1, 1,\ldots, 1)^T$ is one-column identity matrix and S is a symmetric $m_x m$ matrix with elements $s_{ij} = |(X_i - \mu_0)^T(X_j - \mu_0)|$.

3. Determining separated hyper-plane

The interpretation of $L(W, \lambda)$ and its lower bound $L_1(\lambda)$ and upper bound $L_2(\lambda)$ is shown as follows:



3. Determining separated hyper-plane

- By setting derivative of $L(W, \alpha)$ with regard to W to 0, we have:

$$W - \sum_{i=1}^{m} \lambda_i \frac{W^T(X_i - \mu_0)}{|W^T(X_i - \mu_0)|}(X_i - \mu_0) = 0 \quad (1)$$

- Maximizing $L(W, \lambda)$ is equivalent to maximizing $L_1(\lambda)$ and $L_2(\lambda)$. Let λ_1^* and λ_2^* be maximum points of $L_1(\lambda)$ and $L_2(\lambda)$, respectively:

$$\lambda_1^* = \underset{\lambda}{\operatorname{argmax}} L_1(\lambda) = \underset{\lambda}{\operatorname{argmax}} \left(\lambda^T I - \frac{3}{2} \lambda^T S \lambda \right)$$

$$\lambda_2^* = \underset{\lambda}{\operatorname{argmax}} L_2(\lambda) = \underset{\lambda}{\operatorname{argmax}} \left(\lambda^T I - \frac{1}{2} \lambda^T S \lambda \right)$$

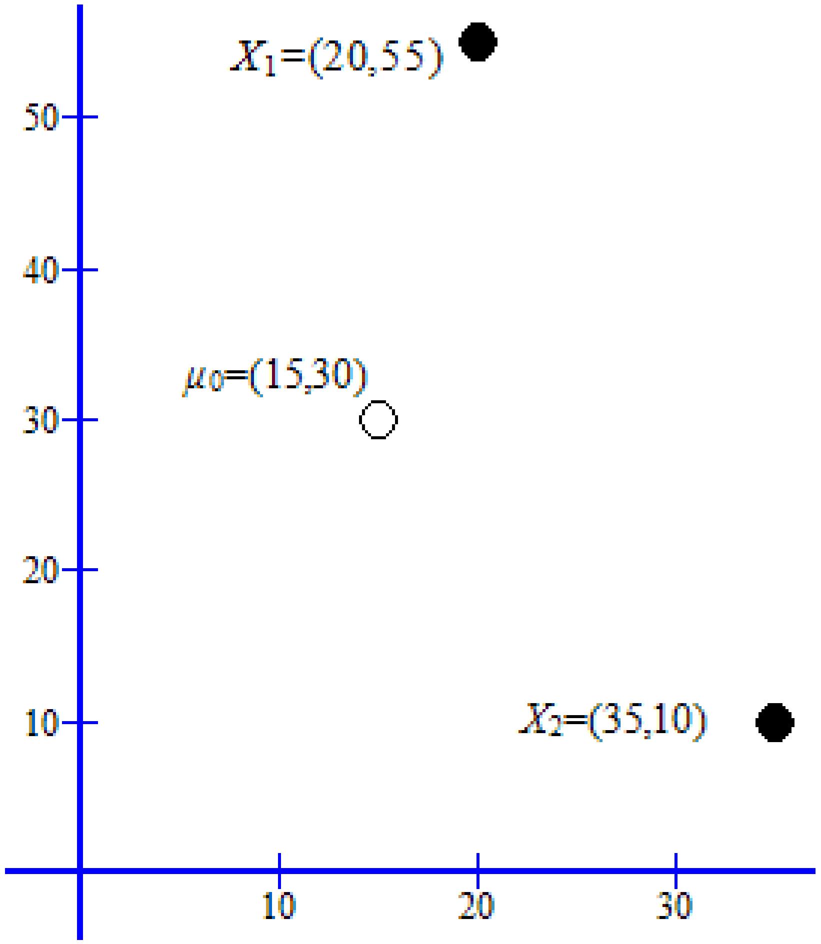
- The maximum point λ^* of $L(W, \lambda)$ is approximated as the mean $\lambda^* = (\lambda_1^* + \lambda_2^*)/2$. We have:

$$W - \sum_{i=1}^{m} \lambda_i^* \frac{W^T(X_i - \mu_0)}{|W^T(X_i - \mu_0)|}(X_i - \mu_0) = 0 \quad (2)$$

- If equation 2 has only one solution W^*, then W^* is normal vector of separated hyper-plane. Otherwise equation 2 has $k > 1$ solutions $W_1^*, W_2^*, \ldots, W_k^*$, then W^* of separated hyper-plane is the addition of these solutions: $W^* = \sum_{i=1}^{k} W_i^*$

4. A case study of separated hyper-plane test

Given a sample including two data points X_1={20, 55}, X_2={35, 10} and a null median $\tilde{\mu}_0 = \{15,30\}$, it is necessary to perform a sign test with null hypothesis H_0: $\tilde{\mu} = \tilde{\mu}_0$ in flavor of alternative hypothesis H_1: $\tilde{\mu} \neq \tilde{\mu}_0$

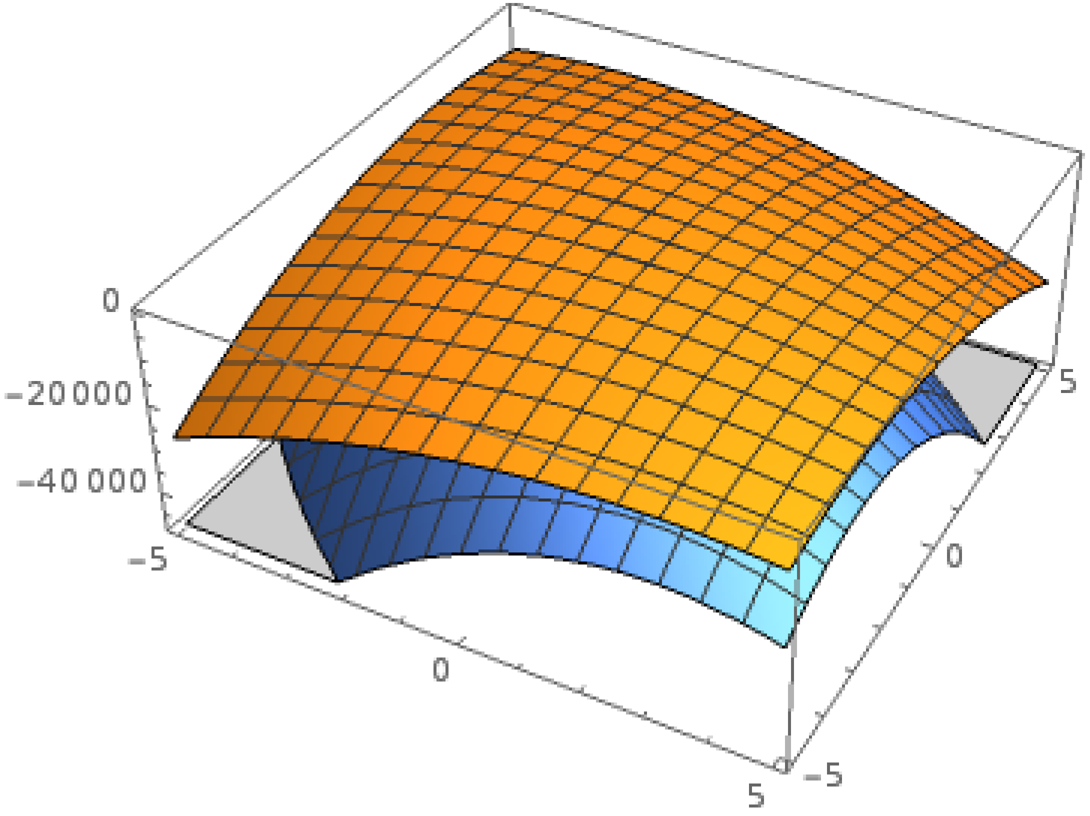


4. A case study of separated hyper-plane test

Lower bound and upper bound of $L(W, \lambda)$ are:

$$L_1(\lambda) = \lambda_1 - 975\lambda_1^2 + \lambda_2 - 1200\lambda_1\lambda_2 - 1200\lambda_2^2$$
$$L_2(\lambda) = \lambda_1 - 325\lambda_1^2 + \lambda_2 - 400\lambda_1\lambda_2 - 400\lambda_2^2$$

Next figure shows 3-dimension graphs of $L_1(\lambda)$ $L_2(\lambda)$.



4. A case study of separated hyper-plane test

- The $L_1(\lambda)$ and $L_2(\lambda)$ get maximal at $\left(\frac{1}{2700}, \frac{1}{4320}\right)$ and $\left(\frac{1}{900}, \frac{1}{1440}\right)$, respectively. the maximum point λ^* is the average as follows:

$$\lambda^* = \frac{1}{2}\left(\left(\frac{1}{2700}, \frac{1}{4320}\right) + \left(\frac{1}{900}, \frac{1}{1440}\right)\right) = \left(\frac{1}{1350}, \frac{1}{2160}\right)$$

- Solving following equation to get W^*:

$$W - \sum_{i=1}^{m} \lambda_i^* \frac{W^T(X_i - \mu_0)}{|W^T(X_i - \mu_0)|}(X_i - \mu_0) = 0 \Leftrightarrow \begin{cases} w_1 - \dfrac{w_1 - w_2}{108|w_1 - w_2|} - \dfrac{w_1 + 5w_2}{270|w_1 + 5w_2|} = 0 \\ w_2 + \dfrac{w_1 - w_2}{108|w_1 - w_2|} - \dfrac{w_1 + 5w_2}{270|w_1 - 5w_2|} = 0 \end{cases}$$

- The normal vector $W^* = \begin{pmatrix} w_1 = -0.00555556 \\ w_2 = 0.0277778 \end{pmatrix}$ is solution of the above equation.

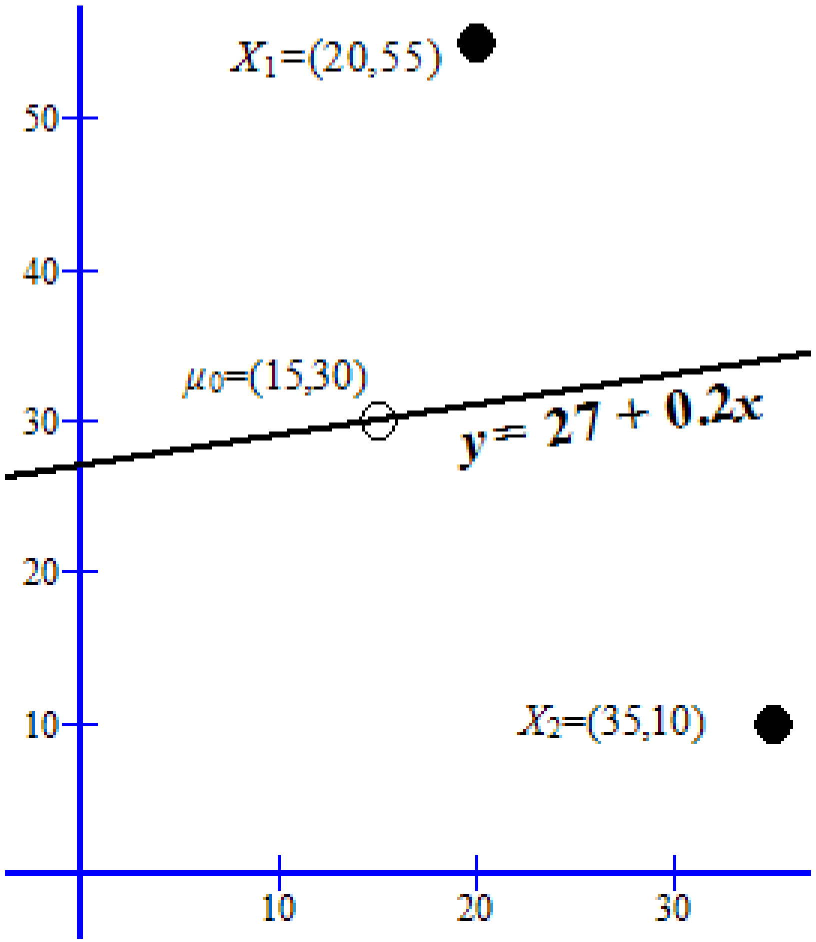
4. A case study of separated hyper-plane test

- Given $W^* = \begin{pmatrix} w_1 = -0.00555556 \\ w_2 = 0.0277778 \end{pmatrix}$, the separated hyper-plane is determined as follows:

  $W^{*T}(X - \mu_0) = 0$

  $\Leftrightarrow (-0.00555556, 0.0277778)\begin{pmatrix} x - 15 \\ y - 30 \end{pmatrix} = 0$

  $\Leftrightarrow y = 27 + 0.2x$
- Next figure 4 shows such separated hyper-plane:



4. A case study of separated hyper-plane test

- Due to:

$$W^T X_1 = -0.00555556 * 20 + 0.0277778 * 55 \approx 0.67 > 0$$
$$W^T X_2 = -0.00555556 * 35 + 0.0277778 * 10 \approx -0.67 < 0$$

- There are 1 positive sign and 1 negative sign. According to sign test, that the number of plus signs is equal to the number of minus signs leads to fail to reject the null hypothesis H_0: μ=(15,30).

5. Conclusions

- The main idea of this paper is to find out a separated hyper-plane, which aims to count the number of observations which fall in left side or right side of such hyper-plane.
- The method to find out hyper-plane is a variant of support vector machine (SVM) [2] method except that hyper-plane contains the null hypothesis. SVM assumes that data points are labeled with classes {1 and –1} but observations in case of hypothesis testing are not classed. This research solves this problem by two-step process of separation:
  1. Firstly, the label of observations is specified indirectly via the scalar product between these observations and the normal vector of separated hyper-plane.
  2. Secondly, Lagrange multipliers are determined via arithmetic transformations when the goal of such transformations is to eliminate normal-vector-dependency from Lagrange function and find out the lower bound and supper bound of Lagrange function. After that the best-separated hyper-plane is the intermediate one among many possible hyper-planes.

Thank you for attention

References

1. R. E. Walpole, R. H. Myers, S. L. Myers and K. Ye, Probability & Statistics for Engineers & Scientists, 9th ed., D. Lynch, Ed., Boston, Massachusetts: Pearson Education, Inc., 2012, p. 816.
2. M. Law, “A Simple Introduction to Support Vector Machines,” 2006.
3. S. Boyd and L. Vandenberghe, Convex Optimization, New York, NY: Cambridge University Press, 2009, p. 716.
4. J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," Microsoft Research, 1998.
5. R. L. Burden and D. J. Faires, Numerical Analysis, 9th Edition ed., M. Julet, Ed., Brooks/Cole Cengage Learning, 2011, p. 872.
6. I. Johansen, Graph software, GNU General Public License, 2012.
7. Wolfram, Mathematica, vol. 10, Wolfram Research.