**A heuristic approach to determine the hyper-plane that contains specified vector and separates vector space into two half-spaces**

Loc Nguyen

Vietnam Institute of Mathematics

**Abstract**

There is the problem that given *n-*dimension vector space *X* and given a specified vector *x0* how to specify the hyper-plane which separates such vector space as discriminately as possible and such hyper-plane must contain specified vector *x0*. In order to solve this problem, this research creates a space-separation equation whose solution is such hyper-plane by mathematical proof and proposes a heuristic algorithm to solve such equation. This algorithm is a variant of Newton’s method and its goal is to reduce the complexity of solving equation and it is always to find out the hyper-plane or assert that such hyper-plane is not existent.

**Keywords**: separated hyper-plane, half-spaces, Newton’s method.

Bài toán của nghiên cứu này được nảy sinh từ việc làm thế nào để xác định một siêu phẳng chứa một điểm *x0* và tách càng phân biệt càng tốt một không gian vector *X* cho trước. Nghiên cứu đã giải quyết bài toán này bằng cách xây dựng một (hệ) phương trình toán học chia tách không gian vector, đồng thời đề xuất một thuật toán thông minh để giải phương trình này. Thuật toán thông minh là một biến thể của phương pháp Newton nhằm giảm độ phức tạp tính toán khi giải phương trình và bảo đảm phương trình luôn giải được hoặc xác định rằng siêu phẳng không tồn tại.

**1. Equations of separating hyper-plane**

This research is originated from the problem that given a sample vector space *X =* {*x*1, *x*2,…, *xm*} where *xi* (s) are *n-*dimension sample vectors, *xi =* {*xi*1, *xi*2,…, *xin*} and given a specified vector *x*0 = {*xi*1, *xi*2,…, *xin*} how to specify the hyper-plane which separates such vector space as discriminately as possible and such hyper-plane must contain specified vector *x*0. This research is very similar to data clustering and classification but the constraint vector *x*0 makes applications of this research different from ones of other methods when constraint vector *x*0 is the link point to split data into two parts. So the optimal hyper-plane *W* satisfies two following propositions:

1. *W* contains specified vector *x*0.
2. *W* splits vector space *X* into two half-spaces as separately as possible.

The first proposition is by following equation that is the equation of hyper-plane.

*wT*(*xi – x*0) = 0

Where *w* is the normal vector of hyper-plane and *T* denotes transposition operator. The task of determining optimal hyper-plane is identical to find out its normal vector *w.* The second proposition is equivalent to maximizing the distance between two margins of two half-spaces. This proposition is similar to the methodology of support vector machine (SVM) [2] method. These margins between two half-spaces are restricted by two parallel hyper-planes that are specified by following equations:

Two half-spaces are separated as much as possible if and only if following condition is satisfied:

Because the distance between two margins is , the optimal hyper-plane will maximize . It means that the optimal hyper-plane minimizes with constraint . According to Lagrange dual theorem [1], *w* is extreme point of following Lagrange function:

Where *λ =* (*λ*1, *λ*2,…, *λm*)*T* represents a set of Lagrange multipliers and . Suppose *w* minimizes then it is the solution of following equation when we set the derivative of *L*1(*w*, *α*) with regard to *W* to 0.

Substituting (2) into equation (1), we have:

There are two approaches to find out the optimal normal vector *w\** which is the solution of both equation (2) and (3). The first approach is to determine the inferior and supreme of *L*(*w*, *λ*) so that the extreme point of *L*(*w*, *λ*) is calculated based on these inferior and supreme, which in turn the optimal normal vector *w\** is determined according to extreme point. The second approach is described in this research where both equations (2) and (3) constitute a set of equations. *L*(*w, λ*) is re-written in matrix notation.

Where *G*(*w*) is symmetric matrix whose each element is the function of *w* and 𝟙 = (1, 1,…, 1)*T* is identity vector. Minimizing is identical to maximizing *L*(*w*, *λ*) with regard to *λ*. Suppose *λ\** is the extreme point of *L*(*w*, *λ*) when *L*(*w*, *λ*) is considered as the function of *λ*, we set the derivative of *L*(w*, λ*) with regard to *λ* to 0 so as to find out *λ\**.

Combining (2) and (4), we have the set of equations:

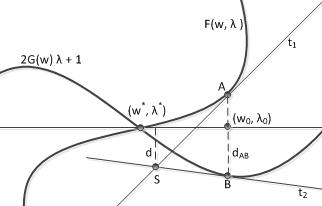
The set of equations (5) is called separation equation in brief. Note that *F*(*w, λ*) is the *w-λ*-dependent vector and *G*(*w*) is the *w-*dependent matrix. The pair of best normal vector *w\** and Lagrange multipliers *λ\** is the solution of separation equation (5). A heuristic method proposed to solve thisseparation equation is described in next section.

**2. Heuristic algorithm to solve separation equation**

As aforementioned the best normal vector *w\** and Lagrange multipliers *λ\** are the solutions of separation equation (5). In other words, two curved surfaces *F*(*w*, *λ*) = 𝟘 and 2*G*(*w*)*λ* + 𝟙 = 𝟘 specified in (5) are cut together at the plane created by co-ordination axes. Suppose that our co-ordinate system has three axes such as *w*, *λ* and *z* where *z* representsevaluations on *F*(*w*, *λ*) and *2G*(*w*)*λ* + 𝟙. Research proposes a heuristic algorithm to find out the best normal vector *w\** and Lagrange multipliers *λ\**. This algorithm is an extension of Newton’s method [3], which includes two following steps:

1. Given point *P*(*w*0, *λ*0), if *P* is not existent, then *P* is initialized arbitrarily. The tangents of *F*(*w*, *λ*) and 2*G*(*w*)*λ* + 𝟙at point *P*(*w*0, *λ*0) are calculated. Let *t*1 and *t*2 be the tangents of *F*(*w*, *λ*) and 2*G*(*w*)*λ* + 𝟙, respectively. If *t*1 and *t*2 are secant, go to step 2; otherwise repeating choosing another arbitrary point *P*(*w*0, *λ*0) until *t*1 and *t*2 are secant. Note that point *P*(*w*0, *λ*0) must be in the domain of both *F*(*w*, *λ*) and 2*G*(*w*)*λ* + 𝟙. If *t*1 and *t*2 are not secant after *k* iterations, algorithm is terminated with conclusion that equation (5) has no solution.
2. Suppose *S*(*w*, *λ*, *z*) is the intersection point between tangents *t*1 and *t*2. Let *A* and *B* are points belonging to *F*(*w*, *λ*) and 2*G*(*w*)*λ* + 𝟙, respectively so that *A* and *B* have the same abscissa(*w*, *λ*) to *S*. Let *d* be the distance between *S* and abscissa plane created by *w-*axis and *λ-*axis. Let *dAB* be the distance between *A* and *B*. Given pre-defined thresholds *δ* and *ε*, if *d < δ* and *dAB < ε*, then algorithm is stopped and the solution of separation equation is *S* = (*w\**, *λ\**, *0*). Otherwise go back step 1 with *P* = *S*.

This algorithm finds the approximate solutions and the accuracy is based on pre-defined thresholds *δ* and *ε*. If solutions are existent, this algorithm is always convergent. Following figure is the imagination of proposed algorithm.



**4. Conclusion**

The basic idea of this research is to unify Lagrange constraints into an equation so as to solve out the best normal vector of separated hyper-plane. The proposed algorithm solving such equation is a variant of Newton’s method with assumption that the solution of separation equation is approximate to the intersection of surfaces created by Lagrange constraints. This algorithm is creative invention to reduce the complexity of solving multivariate equation but it doesn’t deliver roots as exactly as traditional Newton’s method does. The accuracy is based on pre-defined heuristic thresholds. Moreover arbitrarily initialized points in step 1 affect significantly on the convergence speed; in other words time cost is inversely proportional to quality of initialized points when the quality of initialized points are measured by the distance between them and real solution. In the future, research will propose techniques to enhance the quality of initialized points so as to be as near to real solution as possible.

**Reference**

1. [Boyd, Vandenberghe 2004]. Stephen Boyd, Lieven Vandenberghe. Convex Optimization. Copyright © Cambridge University Press 2004. ISBN 978-0-521-83378-3.
2. [Cristianini, Shawe-Taylor 2000]. N. Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines. Cambridge University Press, 2000.
3. [Wikipedia modified 2013]. Newton’s method - Wikipedia, the free encyclopedia, modified 2013. Website: https://en.wikipedia.org/wiki/Newton%27s\_method.