CENTER FOR
PARALLEL OPTIMIZATION

CANCER DIAGNOSIS VIA LINEAR PROGRAMMING

by

O. L. Mangasarian & W. H. Wolberg

Computer Sciences Technical Report #958

August 1990

Cancer Diagnosis via Linear Programming*

O. L. Mangasarian
Computer Sciences Department
&
W. H. Wolberg
Department of Surgery

University of Wisconsin - Madison

Abstract

This report describes informally a diagnostic system that has been in operation at University of Wisconisn Hospitals for the past 17 months. The system, which has correctly diagnosed 165 out of 166 cases in that period, is based on linear programming and can also be viewed as a neural network.

*The research described in this article was supported by Air Force Office of Scientific Research Grant AFOSR-89-0410 and National Science Foundation Grant CCR-8723091.

Cancer Diagnosis via Linear Programming

O. L. Mangasarian
Computer Sciences Department
&
W. H. Wolberg
Department of Surgery

University of Wisconsin - Madison

For the last seventeen months a computer program generated by linear programming has been in continuous use at University of Wisconsin Hospitals in Madison for the diagnosis of breast cancer. The program has correctly diagnosed 165 out of the 166 cases processed during that period. In this article we give a brief description of this important application of linear programming by describing how this diagnostic program is generated and kept up to date.

The program is based on the multisurface method (MSM) of pattern separation [4] that was proposed by one of us in 1968 and which has been recently [1] shown to be a method for training neural networks with partially pre-assigned weights. Fundamental to MSM is the role played by linear programming as the key algorithm for generating separating planes. These planes constitute a piecewise-linear surface that distinguishes benign from malignant cases in the diagnostic program.

We begin by first describing the medical diagnostic procedure. When a patient is examined for breast cancer, a fine needle aspirate (fna) is obtained from the suspected malignant area. The following 9 quantities are then observed under a microscope and assigned a number between 1 and 10 by the examining physician: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses [8]. Larger numbers generally in-

dicate a greater likelihood of malignancy, although no single measurement by itself can determine whether the sample is benign or malignant. If the fna is determined to be benign the patient is reexamined after 3 months and 1 year, or given the choice of a biopsy. A biopsy, which is a more involved surgical procedure and is 8-times more costly, establishes positively whether the patient has a malignancy or not. A biopsy is performed on all patients with fna's indicating malignancy. Thus the diagnostic problem consists of determining from the nine measurements on the fna whether the patient has a malignancy or not. This diagnosis is confirmed by either reexamination (for benign fna) or by biopsy (for all malignant fna's and some benign fna's). Figure 1 shows a benign fna and Figure 2 shows a malignant fna, both magnified 6250 times. These two samples were borderline cases that caused some difficulty for the examining physician, but were classified correctly without difficulty by the diagnostic program.

We describe now the mathematical pattern separation problem associated with the above diagnostic procedure. Given a training set consisting of 2 disjoint point sets in the 9-dimensional real space $\mathbb{R}^9$ representing confirmed benign and malignant fine needle aspirates, construct a discriminant function that distinguishes between those 2 sets. This discriminant function will then be used for predicting whether a new aspirate is benign or malignant. Let us call these two disjoint point sets B (benign) and M (malignant). If their convex hulls do not intersect, then it is a simple matter to discriminate between them by a single separating plane that is easily generated by solving one linear program [3, 5]. However in general, the convex hulls of B and M do intersect, in which case we have to resort to separation by a piecewise-linear surface which is generated as follows by solving a finite sequence of linear programs. We first construct a pair of parallel planes in $\mathbb{R}^9$ which are as close to each other as possible such that the closed region between them contains the intersection of the convex hulls of the sets B and M, and such that each of the opposite open halfspaces determined by the pair of parallel planes contains points from either B only or M only. The parts of B and M contained in these open halfspaces are therefore separated from each other and are removed from the problem. (For certain uncommon degenerate cases the two open halfspaces may be empty, in which case an antidegeneracy step must be added [4, 5].) The procedure is then repeated for the remaining points lying in the closed region between the planes. In this way an ordered finite number of parallel planes eventually separate the

two sets. These planes constitute a piecewise-linear discriminant function which completely separates the given training set and will be used to predict whether a given new point is benign or malignant by determining in which of the ordered halfspaces it lies.

As mentioned above, the principal tool in our approach is linear programming [2] which is one of the most seminal concepts of modern applied mathematics. However its use in the above problem is somewhat delicate, because of a lurking NP-complete problem for the linearly inseparable case. This is the case because the underlying problem is a nonconvex optimization problem, with the nonconvexity induced by a single nonzeroness constraint of the type: $norm(x) \geq 1$. Depending on the norm employed, the problem may become NP-complete [5, 6]. For example if a 2-norm is employed, the problem is NP-complete whereas if the infinity-norm is employed, the problem can be solved in polynomial time by solving a finite number of linear programs, which is precisely what is done in our approach [5, 1].

We mention briefly the connection between MSM and neural networks [1]. Neural networks [7] have gained wide popularity in a number of fields and especially in artificial intelligence. Fundamentally a neural network can be thought of as a nonlinear map between two spaces. The training procedure consists of determining map parameters so as to fit a given set of data (the training set). The map is then used for prediction on new data (the testing set). In artificial intelligence parlance a neural net is represented by inputs units (that accept data from the domain space $\mathbb{R}^9$ for our problem), weighted arcs from the input units to a number of hidden units (that generate step functions at thresholds to be determined by the training algorithm), and finally weighted arcs from the hidden units to an output unit (that generates an output in the range space {0,1} for our problem). Our MSM approach is a method for training or determining the parameters of this neural net. The weights of the incoming arcs to the hidden units, are the normals to the pairs of parallel planes, the weights of the outgoing arcs are predetermined to produce a precedence relation among the outputs of the hidden units that corresponds to the ordering of the pairs of halfspaces of MSM, and finally the threshold values for the hidden units are given by the (positive or negative) distance of each of the planes from the origin in $\mathbb{R}^9$.

We give now a description of how our diagnostic program was trained

and how it is updated. In January of 1989, 369 points in $\mathbb{R}^9$, (201 benign, 168 malignant) were used by MSM to generate 4 pairs of parallel planes that completely separated the benign from the malignant points. The resulting program was used at University of Wisconsin Hospitals until October 1989 to diagnose 70 new cases, all of which were diagnosed correctly except one. At that time, all the available 439 points (258 benign, 181 malignant) were used to generate 4 new pairs of planes that completely separated the benign from malignant points. Since that time, these new planes have correctly classified all 96 new points (72 benign, 24 malignant) encountered. The new planes will continue to be used as long as no misclassification occurs. If and when a misclassification occurs, the system will then be retrained on all the available data at that time. A database containing all 535 points (330 benign, 205 malignant) is now available by electronic mail from *olvi@cs.wisc.edu.* The database in fact contains only 383 unique points (181 benign, 202 malignant) and 152 duplicates (149 benign, 3 malignant).

We hope that we have conveyed the essence of our approach in this brief and informal article. We have shown that MSM is a powerful tool for pattern separation that has a very important application in medical diagnosis. Other possible applications include finance, oil exploration, decision sciences and inverse problems.

References

1. K. P. Bennett & O. L. Mangasarian: "Neural network training via linear programming", University of Wisconsin Computer Sciences Tech. Rpt. 948, July 1990.

2. G. B. Dantzig: "Linear programming and extensions", Princeton University Press, Princeton, New Jersey 1963.

3. O. L. Mangasarian: "Linear and nonlinear separation of patterns by linear programming", Operations Research 13, 1965, 444-452.

4. O.L. Mangasarian: "Multi-surface method of pattern separation", IEEE Transactions on Information Theory, IT-14, 1968, 801-807.

5. O. L. Mangasarian, R. Setiono & W. H. Wolberg: "Pattern recognition via linear programming: Theory and application to medical

diagnosis", University of Wisconsin Tech. Rpt. 878, September 1989, to appear, Proceedings of SIAM Workshop on Large-Scale Numerical Optimization, Cornell University, Ithaca, New York, October 19-20, 1989.

6. O. L. Mangasarian & T.-H. Shiau: "A variable-complexity norm maximization problem", SIAM Journal on Algebraic and Discrete Methods 7, 1986, 455-461.

7. D. E. Rumelhart & J. L. McClelland (eds): "Parallel distributed processing", Volume I, M.I.T. Press, Cambridge, Massachusetts 1986.

8. W. H. Wolberg & O. L. Mangasarian: "Multisurface method of pattern separation applied to breast cytology diagnosis", April 1989, submitted, Proceedings of the National Academy of Sciences U.S.A.