The Delaunay Triangulation and Function Learning

STEPHEN M. OMOHUNDRO

International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, California 94704
Phone: 415-643-9153
Internet: om@icsi.berkeley.edu
Date, 1989

Abstract. In this report we consider the use of the Delaunay triangulation for learning smooth nonlinear functions with bounded second derivatives from sets of random input output pairs. We show that if interpolation is implemented by piecewise-linear approximation over a triangulation of the input samples, then the Delaunay triangulation has a smaller worst case error at each point than any other triangulation. The argument is based on a nice connection between the Delaunay criterion and quadratic error functions. The argument also allows us to give bounds on the average number of samples needed for a given level of approximation.

Introduction

Elsewhere we have described the importance to robotics, machine vision, speech processing and graphics of algorithms for learning smooth relationships between variables from examples [Omohundro 1987, 1988, 1989a]. In this report we consider the task of learning nonlinear mappings from $\Re^k$ to $\Re^m$. In fact we only explicitly discuss mappings to $\Re$, but the results extend straightforwardly to $\Re^m$. In many applications such mappings do not vary their slope too quickly. We model this characteristic by considering the class of functions which have a bounded second derivative in each direction. A natural approach to high-dimensional interpolation is to triangulate the input points and use linear interpolation within each triangle. This approximation has the virtue of being continuous and as we shall see, amenable to *a priori* bounds on the error of approximation. It still leaves open the question of which triangulation to choose. In this report we show that if one is concerned with choosing a single triangulation which will do well over all functions with a second-derivative bound, then a particular triangulation, the Delaunay triangulation is an optimal choice.

This choice also has considerable computational advantages which we discuss elsewhere [Omohundro, 1989a]. Triangulations typically consist of a large number of facets in high dimensional spaces. If one needs to precompute all of these, the computational burden can become unreasonable in high dimensions. The simplices in the Delaunay triangulation, however, may be determined by a local criterion. The simplex containing a point may be computed on the fly as requests are received. Arguments similar to those in [Friedman, *et. al.,* 1977] and [Devroye, 1986] for fast nearest neighbor queries may be applied to show that for points drawn from a smooth underlying distribution, the Delaunay simplex containing a given point may be found in a time which is asymptotically only logarithmic in the number of samples. The size of the required data structure is only linear in the number of samples.

This report begins with precise definitions of the needed concepts. The second section presents the main theorem and its proof through a series of lemmas. The final section uses one of the lemmas to give *a priori* bounds on the number of samples needed for a given level of approximation.

Basic Concepts

We denote a k-dimensional Euclidean space by $\Re^k$. Points in this space are k-tuples of real numbers $(x_1, \ldots, x_k)$. Two points determine a unique line segment which joins them. A set $S \subset \Re^k$ is *convex* if it contains all of the line segments determined by pairs of its points. The *convex hull* of a set is the smallest convex set containing it. The convex hull of $j+1$ points $p_1, \ldots, p_{j+1}$ in $\Re^k$, with $j \le k$, is a higher dimensional analog of a triangle or tetrahedron. If the dimension of the hull is $j-1$, we call it a j-simplex and say that the points are *independent.* Some authors refer to the set of points themselves or the interior of the hull as the simplex. The points p_i are the *vertices* of the simplex and any subset of them of size $l+1$ determines an l-simplex called an *l-face* of the original simplex. If $l < j$ we call it a *proper face*.

We say that a set of points in $\Re^k$ is *in general position* if each subset of fewer than $k+1$ points is independent. A set of points drawn from a smooth probability distribution will be independent with probability 1. The convex hull H of a collection of more than $k+1$ points p_i in general position may be decomposed into simplices with the points as vertices. We say that a set of k-simplices with vertices in the p_i is a *triangulation* of H over the p_i, if their union equals H and if any two intersect in only proper faces. Often one extends such a triangulation to the whole of $\Re^k$ by in essence including the point at infinity.

The set of points in $\Re^k$ which are a fixed distance from a point p form a $(k-1)$ *-sphere* centered at p and the sphere together with its interior define a *k-ball*. The sphere determined by a set of $j \le k+1$ independent points is the unique sphere of smallest radius which includes the points (note that the sphere itself must contain the points, not its interior). The radius of this sphere may be explicitly written in terms of the coordinates of the points through the use of Cayley-Menger determinants [Berger, 1987]. Given a set of $l > k$ points p_i in $\Re^k$, we say that a j-simplex determined by $j+1 \le k+1$ of the points is a *Delaunay simplex* if the interior of the sphere that the simplex determines doesn't contain any of the other p_i. The set of all Delaunay k-simplices forms a triangulation of the convex hull of the p_i called the *Delaunay triangulation* [Preparata and Shamos, 1985].

Optimality of the Delaunay Triangulation

In this section we study piecewise-linear approximations of real-valued functions on $\Re^k$. We are given the value of the function at the N sample points p_i in general position in $\Re^k$ for $1 \le i \le N$ and would like to approximate it at other points. The strategy we will use is to choose a triangulation with the given points as vertices and to linearly interpolate the function within each simplex of the triangulation. We will only consider the approximation inside the convex hull of the p's (so as to study interpolation rather than extrapolation). The approximating functions are then piecewise linear and continuous throughout the domain of interest.

Different triangulations will give rise to different approximations and in this section we prove a theorem to motivate the use of the Delaunay triangulation for this application. Because we are using piecewise-linear approximations, we will consider functions whose second derivative is bounded. Let us denote by F_C the set of all real-valued functions on $\Re^k$ such that the absolute value of the second derivative along any straight line is less than the constant $2C$. These functions are analogous to the submanifolds of bounded curvature that we encounter in the surface learning case. We will show that within this class, the Delaunay triangulation gives rise to a piecewise linear approximation with the smallest maximum error at each point over the functions in F_C.

Theorem 1. Let N sample points $p_i \in \Re^k$ in general position be given and denote their convex hull by $H \subset \Re^k$. Let $y \in H$ be a test point of interest. For each function $f \in F_C$, let f_D be the piecewise linear approximation to f defined on H over the Delaunay triangulation D of the vertices p_i and let f_T be the piecewise linear approximation over any other triangulation T of these vertices. For any such triangulation T and any point $y \in H$ we have:

$$max_{f \in F_C} |f_D(y) - f(y)| \le max_{f \in F_C} |f_T(y) - f(y)| . \qquad \textbf{(EQ 1)}$$

Thus the maximum error possible with the Delaunay triangulation is less than with any other triangulation.

We will prove this theorem using a succession of lemmas.

Lemma 1. Let g be a real-valued function on the interval $[0, L] \subset \Re$ with a bounded second derivative $|g''(x)| \le 2C$ and bounds on the values at the endpoints $|g(0)| \le a$ and $|g(L)| \le b$, where $a, b \ge 0$. Then at every point x in the interval:

$$|g(x)| \le C(Lx - x^2) + \frac{b-a}{L}x + a. \qquad \textbf{(EQ 2)}$$

Notice that at any fixed point x, the bound on the magnitude of $g(x)$ is an increasing function of C, L, a, and b.

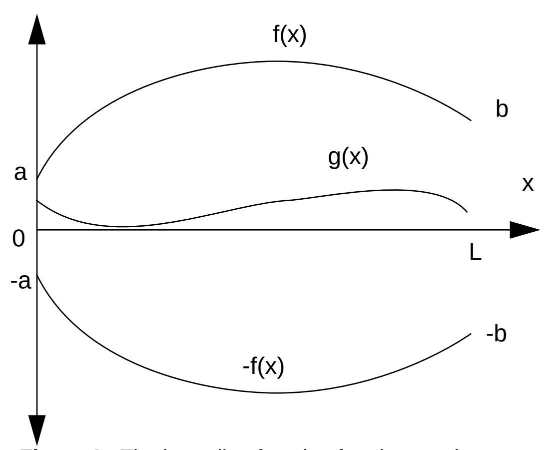

Figure 1. The bounding function f and example function g.

Proof. Let us denote the comparison function on the right hand side of the inequality by $f(x)$. Notice that $f(x) \ge 0$ and $f''(x) = -2C$ for $x \in [0, L]$ and that $f(0) = a$ and $f(L) = b$. We begin by showing that $g(x) \le f(x)$ under only the assumption that $g''(x) \ge -2C$, $g(0) \le a$, and $g(L) \le b$. The other half, i.e. that $-g(x) \le f(x)$, follows from an identical argument using the rest of the conditions. Let us consider the difference function $d(x) \equiv g(x) - f(x)$. By the hypotheses, $d(0) \le 0$, $d(L) \le 0$, and $d''(x) \ge 0$ for all $x \in [0, L]$. First let us show that $d'(x)$ is non-decreasing. If $x_2 \ge x_1$ then

$$d'(x_2) = \int_{x_1}^{x_2} d''(x)\,dx + d'(x_1) \ge d'(x_1). \qquad \textbf{(EQ 3)}$$

If g is a counter-example to the lemma and $y \in [0, L]$ is a point at which $g(y) > f(y)$, then $d(y) > 0$. We will show that this cannot happen by considering $d'(y)$. If $d'(y) \ge 0$ then because $d'(x)$ is non-decreasing, we have $d'(x) \ge 0$ for all $x \in [y, L]$. But this implies that

$$d(L) = \int_{y}^{L} d'(x)\,dx + d(y) > 0 \qquad \textbf{(EQ 4)}$$

contrary to the condition that $d(L) \le 0$. Similarly, if $d'(y) \le 0$, then $d'(x) \le 0$ for all $x \in [0, y]$ since $d'(x)$ is non-decreasing. But then

$$d(0) = -\int_{0}^{y} d'(x)\,dx + d(y) > 0, \qquad \textbf{(EQ 5)}$$

again contrary to assumption. Q.E.D.

Lemma 2. If g is a real-valued function on the interval $[0, L] \subset \Re$ which has a bounded second derivative $|g''(x)| \le 2C$, which vanishes at the first endpoint $g(0) = 0$, and whose first derivative vanishes at the second endpoint $g'(L) = 0$ then

$$|g(x)| \le C(2Lx - x^2) \qquad \textbf{(EQ 6)}$$

at every point x in the interval. The maximum possible absolute value occurs at the right endpoint and is CL^2. Notice that this error bound increases in magnitude with C and L.

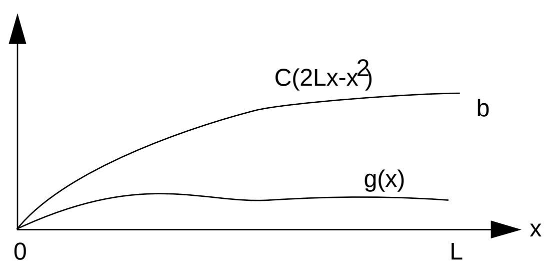

Figure 2. Bounds when g(x)=0 and g'(L)=0.

Proof. As above, we need only consider g's with $g''(x) \geq -2C$. We define $d(x) \equiv g(x) - C(2Lx - x^2)$ giving: $d(0) = 0$, $d'(L) = 0$, and $d''(x) \geq 0$ and we need to show that $d \leq 0$. As above, $d'(x)$ is non-decreasing and so $d'(L) = 0$ implies that $d'(x) \leq 0$ for all $x \in [0, L]$. This immediately shows that

$$d(y) = d(0) + \int_0^y d'(x)\,dx \leq 0 \tag{EQ 7}$$

as desired. Q. E. D.

Lemma 3. Let S be an arbitrary *j*-simplex in $\Re^k$. For every point x in the interior of S, there is a vertex of S such that the distance from x to the vertex is less than or equal to the radius R of the sphere determined by the vertices of S.

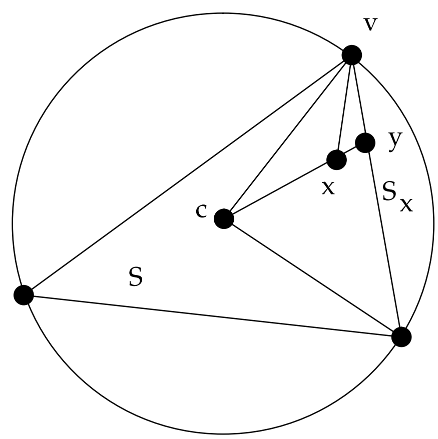

Figure 3. A simplex, whose sphere is centered at c, the point x, and the desired vertex v.

Proof. We prove the lemma by induction on *j*. If *j=2*, and the points are a distance *d* apart, then the sphere they determine is centered at their midpoint and has radius *d/2.* The simplex S is a segment and any point in it is within *d/2* of the closer end. Now let us assume that the lemma is true for all simplices with fewer than *j* vertices and show that it is true for S with *j* vertices. Let the sphere determined by S_j be centered at c and of radius R. Each *j-1* face of S together with c determines a *j* simplex with all edges leaving c of length R. The union of all these simplices contains S. If x is the point of interest, let S_x be a *j-1* face of S such that the simplex determined by $S_x \cup c$ contains x. Let y be the intersection of the line determined by c and x with S_x. The radius R_x of the sphere determined by S_x must satisfy $R_x \leq R$ since S_x is a subset of S. By induction, there is a vertex v of S_x such that the distance from y to v is less than or equal to R_x and therefore R. Since both y and c are a distance less than or equal to R from v, so is every point on the line segment between them (the ball of radius R centered at v is convex). But x is on this line segment, so the distance from v to x is less than or equal to the radius R. Q. E. D.

Lemma 4. (a) Given any *j*-simplex *S* with vertices $\{p_1, \ldots, p_{j+1}\}$ in $\Re^k$, the functions in F_C with the largest absolute error over *S* in an approximation by linear interpolation over the values at the vertices are

$$f(x_1, \ldots, x_k) = A + \sum_{i=1}^{N} B_i x_i - C \sum_{i=1}^{k} x_i^2 \qquad \textbf{(EQ 8)}$$

where *A* and the B_i are arbitrary real constants. Furthermore, the error made in such an approximation of *f* at any point in the simplex is greater than or equal to the error made in approximating any other function in F_C at that point. (b) If *R* is the radius of the sphere defined by *S*, then the magnitude of the error in approximating any function in F_C in the interior of *S* is bounded above by CR^2.

Proof. (a) Let *e(x)* be the error in approximating *f* by its affine approximation. Because we interpolate through the values at the vertices p_i of *S*, *e* must vanish there: $e(p_i) = 0$. Since *e* differs from *f* only by linear and constant terms, we see also that $e''(t) = f''(t) = -2C$ along any straight segment in *S* parameterized by the distance *t* along the segment. If we consider an arbitrary segment in *S* and take *a* and *b* to be the values of *e* on its endpoints, then the restriction of *e* to the segment is exactly the bounding function defined in lemma 1. We prove the present lemma by induction on the dimension *j* of the simplex. If *j=1*, we need only consider the segment defined by the simplex itself as the domain for the lemma above with *a=0* and *b=0* to conclude that the absolute error for any other function must be less than or equal to *e(x)* at any point $x \in S$. Let us now inductively assume the lemma for simplices of dimension smaller than *j* to prove it for an arbitrary *j*-simplex *S*. If the point of interest $x \in S$ is in the boundary of *S*, then we just use the inductive result on the face containing *x*. We may do this because the restriction of a function in F_C to a face is in F_C for that face. If *x* is in the interior, choose a vertex *v* of *S* and let *y* be the intersection of the straight line defined by *v* and *x* and the *j*-1-face opposite *v*. Applying the inductive result to that face, we obtain that the error for any function at *y* has magnitude less than or equal to *e(y)*. To obtain the result for *x*, apply lemma 1 to the segment joining *v* and *y* with *a=0* and *b=e(y)*. Because *e* is equal to the bounding function of lemma 1 on this segment, the magnitude of the error in approximating any function in F_C at *x* must be less than or equal to *e(x)*.

(b) We prove this part of the lemma inductively as well. If *j=1*, then by lemma 1 with *a=b=0*, the maximum absolute error is at the center of the segment, a distance *R* from each vertex, and has a value bounded by CR^2. Otherwise we assume the lemma for simplices with dimension less than *j* and prove it for *S* of dimension *j*. For a given choice of function in F_C, let *x* be the point in *S* where the magnitude of the approximation error is maximal. If *x* is in the boundary of *S*, we use the fact that the sphere determined by a face has radius less than or equal to the sphere determined by *S* and the inductive hypothesis to immediately conclude the result. If *x* is in the interior, then because it is the maximum, the derivative of the error in any direction must vanish. By lemma 3, if *R* is the radius of the sphere defined by *S*, there is a vertex *v* within *R* of *x*. Consider the straight segment joining *v* to *x* of length less than R. The error vanishes at v and its first derivative vanishes at x. By lemma 2 the magnitude of the error must be less than or equal to CR^2. Q.E.D.

Lemma 5. If the sphere determined by a simplex S strictly contains a point P, then it also

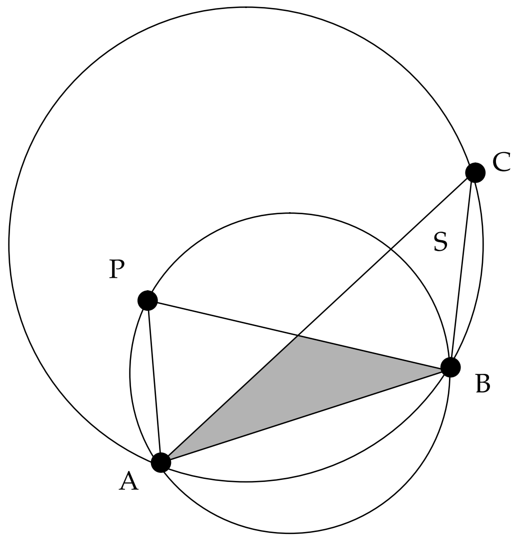

Figure 4. The sphere determined by the simplex ABC contains the point P.

contains each of the simplices determined by P and a face of S. Each point of S is contained in one of these simplices and the worst case error obtained from interpolation on that simplex is less than or equal to the error obtained by using S itself. Because the Delaunay simplices are the only ones whose spheres do not contain other sample points, this shows that at every point their worst case error is less than or equal to that of any other triangulation, and so proving the theorem as desired.

Proof. That S's sphere contains the simplices determined by P and faces of S follows from the convexity of balls and the fact that each point in such a simplex lies on a line segment joining P and a point of the face. To show that each point of S is contained in one of these simplices, we need only extend the line joining it and P until it hits a face of S. The point will be contained in the simplex determined by that face and P. To see that the worst case error functions:

$$f(x_1, \ldots, x_k) = A + \sum_{i=1}^{N} B_i x_i - C \sum_{i=1}^{k} x_i^2 \qquad \text{(EQ 9)}$$

$$\text{(EQ 10)}$$

determined in Lemma 4 are better on these new simplices than on S, we need only consider the difference between them. Let us denote the worst error function on S by f_S and on the new simplex by f_N. We will show that the difference $f_S - f_N$ is greater than or equal to 0 on the new simplex. Notice that the quadratic piece of these two functions is identical and so the difference is an affine function. This affine function vanishes on the vertices of the face of S used in constructing the new simplex and so actually vanishes in the whole hyperplane containing this face. Affine functions are positive on one side of the hyperplane they vanish on and negative on the other. f_N vanishes at P, while f_S is positive there because P is inside its spherical zero set. The difference function is therefore positive at P and so the new simplex lies on the positive side of the zero difference hyper-plane. Q. E. D.

Number of Samples for Given Error

In this section we use the second part of lemma 4 to bound the average number of samples we need to see in order to achieve a given level of approximation. For simplicity, we assume that the samples are drawn from the uniform distribution on the unit cube. For non-uniform, non-vanishing distributions, a similar expression in terms of the minimum value of the probability distribution may be easily obtained. More refined non-uniform estimates may be obtained using the methods in [Devroye, 1986].

Theorem 2. If samples are drawn uniformly from the unit cube in $\Re^k$, then the absolute error made in approximating any function in F_C by the Delaunay method will be less than ε at each point which is at least $k\sqrt{\varepsilon/C}$ from the boundary of the cube after an average of

$$\left(\frac{C}{k\varepsilon}\right)^{k/2}\left(\frac{k}{2}\log\frac{C}{k\varepsilon}+1\right) \qquad \textbf{(EQ 11)}$$

samples have been seen.

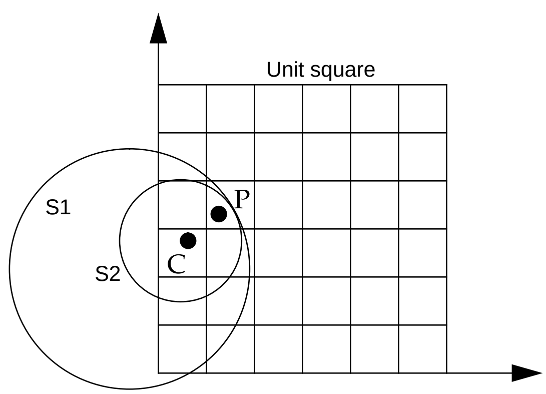

Figure 5. The point P is $\sqrt{2}s$ away from the edge of the unit square and is contained in circle S1. S1 contains S2 which also contains P and has a center C which is inside the unit square. The small cube containing C is entirely inside S2 and so S1.

Proof. Consider a tessellation of $\Re^k$ by cubes of side s. Any ball of radius $R \geq \sqrt{k}s$ must entirely contain the cube which contains its center because a cube's diagonal is of length $\sqrt{k}s$ and is the greatest distance between any two of its points. Now let the unit cube be evenly tessellated by cubes of side s. Consider a point at least $\sqrt{k}s$ away from the edge of the unit cube. Any sphere of radius $R \geq \sqrt{k}s$ which contains this point must also contain a sphere of radius $R = \sqrt{k}s$ which contains the point. Because this sphere's center is in the unit cube, it must completely contain one of the tessellation cubes and therefore so must the original sphere. This shows that if we draw a sample from each tessellation cube, then no Delaunay sphere which contains the test point will have a radius greater than $\sqrt{k}s$. From lemma 4, the error we make at the test point will be less than CR^2. This will be less than ε if $R < \sqrt{\varepsilon / C}$. Thus if we have at least one sample from each cube of side $s = \sqrt{(k\varepsilon)/C}$, the maximum error will be less than ε. There are $(C/(k\varepsilon))^{k/2}$ such cubes.The well-known coupon collector result says that the average time to collect randomly chosen coupons from N people is less than $N(\log N + 1)$. On average then, we will achieve the desired error bound with the stated number of samples. Q.E.D.

Bibliography

Marcel Berger, *Geometry I and II*, (Springer-Verlag, Berlin 1987).

Luc Devroye, *Lecture Notes on Bucket Algorithms,* (Birkhauser, Boston 1986).

Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," ACM Transactions on Mathematical Software}, 3:3 (1977) 209-226.

Stephen M. Omohundro, "Efficient Algorithms with Neural Network Behavior," Complex Systems, 1 (1987) 273-347.

Stephen M. Omohundro, "Foundations of Geometric Learning," University of Illinois Department of Computer Science Technical Report No. UIUCDCS-R-88-1408 (1988).

Stephen M. Omohundro, "Geometric Learning Algorithms", Proceedings of the Los Alamos Conference on Emergent Computation, (1989a). (ICSI Technical Report No. TR-89-041)

Stephen M. Omohundro, "Five Balltree Construction Algorithms", ICSI Technical Report No. TR-89-063 (1989b).

Franco P. Preparata and Michael Ian Shamos, *Computational Geometry, An Introduction*, (Springer-Verlag, 1985).