CONSTRAINED MONOTONE REGRESSION OF ROC CURVES AND HISTOGRAMS USING SPLINES AND POLYNOMIALS

Tapas Kanungo, David M. Gay† and Robert M. Haralick

Intelligent Systems Laboratory
Department of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195
{tapas,haralick}@george.ee.washington.edu

† AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974, USA
dmg@research.att.com

ABSTRACT

Receiver operating characteristics (ROC) curves have the property that they start at (0,1) and end at (1,0) and are monotonically decreasing. Furthermore, a parametric representation for the curves is more natural, since ROCs need not be single valued functions: they can start with infinite slope. In this article we show how to fit parametric splines and polynomials to ROC data with the end-point and monotonicity constraints. Spline and polynomial representations provide us a way of computing derivatives at various locations of the ROC curve, which are necessary in order to find the optimal operating points.

Density functions are not monotonic but the cumulative density functions are. Thus in order to fit a spline to a density function, we fit a monotonic spline to the cumulative density function and then take the derivative of the fitted spline function. Just as ROCs have end-point constraints, the density functions have end-point constraints. Furthermore, derivatives of splines are spline functions and can be computed in closed form. Thus smoothing of histograms can also be treated as a constrained monotone regression problem. The algorithms were implementation in a mathematical programming language called AMPL and results on sample data sets are given.

1. INTRODUCTION

In this paper we consider two problems where monotonic curve fitting with endpoint constraints is necessary. The first problem comes about while fitting a function to data points that represent the operating characteristics of a system (see [1, 2, 3]). Since the receiver operating characteristic (ROC) curve is plot of probability of misdetection versus the probability of false alarm, the data points are monotonically decreasing, and the fitted function also needs to be monotonically decreasing. The second problem is that of fitting smooth functions to normalized histograms. Although probability density functions are not monotonic, the cumulative density functions are. So, one can fit a monotonically increasing function to the normalized cumulative histogram and then take the derivative of the fitted function.

Most of the related work is in the statistics literature. Although the algorithm in [4] results in a smooth histogram, it does not give us a *differentiable function*. The work reported by Ramsay [5] is closest to ours; he forces the fitted spline to be monotonic by using integrated basis splines, which are monotonic, and then constraining the regression coefficients to be non-negative.

In section 2 we pose the constrained monotone regression as an optimization problem that can be solved using standard packages. In section 3 we use the method described in section 2 to fit ROC curves. In section 4 we discuss how to convert the histogram fitting problem into a monotone regression problem.

2. CONSTRAINED MONOTONE REGRESSION

Let (y_i, t_i), $i = 0, \ldots, n-1$, be the given data set, where $y_i \geq y_{i+1}$ and $t_i < t_{i+1}$. The problem is to find a function $y(t)$ that is (i) monotonically decreasing within the interval $[t_0, t_{n-1}]$, (ii) attains given values c_a and c_b at the end-point, that is, $y(t_0) = c_a$, and $y(t_{n-1}) = c_b$, and (iii) minimizes the sum of squared residuals. The monotonicity constraint can be achieved by requiring the derivative to be non-positive, that is, $y'(t) \leq 0$. If we required $y'(t) \geq 0$, then we would get a monotonically increasing $y(t)$.

In the next subsection we let $y(t)$ be a spline function and show find the regression coefficients such that $y(t)$ satisfies all the constrains. In the following subsection we use polynomials for $y(t)$. We implemented both algorithms in AMPL, a mathematical modeling language [6].

2.1. Spline regression

In this section we show how to fit splines to data while satisfying the monotonicity and end-point constraints. The discussion here is based on books by de Boor [7] and Dierckx [8].

A function $y(t)$, defined on a finite interval $[a, b]$, is called a spline function of degree $k > 0$, having as *knots* the strictly increasing sequence λ_j, $j = 0, 1, 2, \ldots, m + 2k + 1$, if the following two conditions are satisfied:

1. On each knot interval $[\lambda_j, \lambda_{j+1}]$, $y(t)$ is given by a polynomial of degree k at most.
2. The function $y(t)$ and its derivatives up to order $k-1$ are all continuous on $[a, b]$.

We will assume that the number of knots and the knot locations λ_j are known. The spline, $y(t)$, can be defined as a linear combination of a finite number of basis splines or B-splines. The jth basis spline, $B_{j,k}(t)$, where k is the degree, is defined recursively as follows

$$\begin{aligned} B_{j,0}(t) &= \begin{cases} 1 & \text{if } \lambda_j \le t < \lambda_{j+1}, \\ 0 & \text{otherwise.} \end{cases} \\ B_{j,k}(t) &= \frac{t - \lambda_j}{\lambda_{j+k} - \lambda_j} B_{j,k-1}(t) + \frac{\lambda_{j+k+1} - t}{\lambda_{j+k+1} - \lambda_{j+1}} B_{j+1,k-1}(t). \end{aligned}$$

B-splines have the property that they are non-negative and vanish unless $\lambda_j \le t < \lambda_{j+k+1}$. In particular, for quadratic splines, if $\lambda_j \le t < \lambda_{j+1}$, then $B_{j-2,2}(t), B_{j-1,2}(t)$, and $B_{j,2}(t)$ are the only non-zero B-splines of degree $k = 2$.

Given the knot locations, λ_j, the spline function $y(t)$ can be evaluated at any parameter location t_i as follows. Let λ_j be such that $\lambda_j \le t_i < \lambda_{j+1}$. Then

$$y(t_i) = \sum_{l=j-k}^{j} \alpha_l B_{l,k}(t_i).$$

where α_l, $l = 0, \ldots, m + k$, are spline coefficients. The monotonicity constraints, in the case of splines, can be imposed by using the derivative properties of spline coefficients (see [7, 8] for proofs).

$$\begin{aligned} \text{If } \alpha_l &\le \alpha_{l+1}, \text{ then } y \text{ is monotonically increasing,} \quad (1) \\ \text{If } \alpha_l &\ge \alpha_{l+1}, \text{ then } y \text{ is monotonically decreasing.} \quad (2) \end{aligned}$$

It is important to note that equation (1) is a *sufficient* but not necessary condition for a spline function to be monotonically increasing; similarly, equation (2) is a sufficient but not necessary condition for a monotonically decreasing spline. *Necessary* conditions are known for the monotonicity of cubic splines [9], but we are unaware of any such necessary conditions for splines of higher order. The constrained optimization problem can now be stated as:
Find $\alpha_0, \ldots, \alpha_{m+k}$, to minimize

$$\sum_{i=0}^{n-1} (y_i - y(t_i))^2$$

subject to

$$\begin{aligned} y(t_0) &= c_a \\ y(t_{n-1}) &= c_b \\ \alpha_l &\ge \alpha_{l+1} \text{ for } l = 0, \ldots, m + k - 1, \end{aligned}$$

where c_a and c_b are given constants. This is a constrained optimization problem (least squares with linear constraints) and can be expressed in AMPL [6] and solved by various solvers, such as MINOS [10] or NPSOL [11]. For a survey and comparison optimization techniques and software, see [12]

Finally, the derivative of a spline $y(t)$ is given by (see [7] for proof):

$$y'(t_i) = \sum_{l=j-k+1}^{j} (k-1) \frac{\alpha_l - \alpha_{l-1}}{t_{l+k-1} - t_l} B_{l,k-1}(t_i). \quad (3)$$

2.2. Polynomial fitting

Now we solve the constrained monotone fitting problem again, but this time we use a polynomial representation for $y(t)$:

$$y(t) \;=\; \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_m t^m,$$

where m is the user-specified degree of the polynomial. We can now write the monotonicity constraint as

$$y'(t) = \beta_1 + 2\beta_2 t + \cdots + m\beta_m t^{(m-1)} \le 0.$$

The above equation constrains the derivative $y'(t)$ to be non-negative. One way of enforcing that is to compute the extrema of $y'(t)$ and constrain them instead. To do that, we would need to find the roots of $y''(t)$ and would need to constrain the values of $y'(t)$ at the roots. In our fitting problem, polynomials of degree $8 - 15$ are commonly used; finding roots of polynomials of degree 13 and constraining function values at those roots appears cumbersome. Instead, we will solve the simpler problem of constraining $y'(t)$ only at sampled points. This may give a nonmonotonic fit, but any departure from monotonicity should small and should decrease if more sample points are added. Thus the constrained optimization problem can now be stated as:
Find $\alpha_0, \ldots, \alpha_m$, to minimize

$$\sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{m} (t_i)^j \alpha_j \right)^2$$

subject to

$$\begin{aligned} \sum_{j=0}^{m} (t_0)^j \alpha_j &= c_a \\ \sum_{j=0}^{m} (t_{n-1})^j \alpha_j &= c_b \\ \sum_{j=1}^{m} j \cdot (t_i)^{(j-1)} \cdot \alpha_j &\ge 0 \text{ for } i = 0, \ldots, n-1, \end{aligned}$$

where c_a and c_b are given constants. This is again a least squares problem with linear constraints.

3. ROC CURVE FITTING

We now apply the regression method discussed in the previous section to ROC data. The receiver operating characteristic curve data is an ordered sequence of 2D points (x_i, y_i), $i = 0, \ldots, n-1$, such that (i) $(x_0, y_0) = (0, 1)$, (ii) $(x_{n-1}, y_{n-1}) = (1, 0)$, (iii) $x_i \leq x_{i-1}$, and (iv) $y_i \geq y_{i+1}$. The problem is to fit a parametric curve $(x(t), y(t))$, $0 \leq t \leq 1$, such that $x(t)$ is monotonically increasing, $y(t)$ is monotonically decreasing, $(x(0), y(0)) = (0, 1)$, and $(x(1), y(1)) = (1, 0)$.

For fitting splines, we choose the knot locations λ_j and parameter locations t_i as follows:

$$\begin{aligned} t_i &= \frac{i}{n-1}, i = 0, \ldots, n-1, \\ \lambda_j &= \frac{j-k}{m}, \; j = 0, \ldots, m+2k+1, \end{aligned}$$

where k is the (user-specified) degree of basis splines, and $m+k+1$ is the number of basis splines. With this choice of parameter locations, t_i, and knot locations, λ_j, we see that $t_0 = \lambda_k = 0$ and $t_{n-1} = \lambda_{m+k} = 1$. Now two separate monotonic functions $x(t)$ and $y(t)$ can be fit to (x_i, t_i) and (y_i, t_i), respectively. In this case $x(t)$ needs to be monotonically increasing, $y(t)$ monotonically decreasing, and both have to satisfy the end-point constraints. Note that other parameterizations such as arc length parameterization could have been chosen instead. The optimization problem can be set up as discussed in section 2. A sample spline fit, computed by our method, is shown in figure 1.

For fitting polynomials, we choose the parameter locations

$$t_i = 2\left(\frac{i}{n-1}\right) - 1.$$

We see that, $t_0 = -1$, and $t_{n-1} = 1$, and the end-point constraints now become

$$\begin{aligned} x(t_0) &= x(-1) = 0 \\ x(t_{n-1}) &= x(1) = 1 \\ y(t_0) &= y(-1) = 1 \\ y(t_{n-1}) &= y(1) = 0. \end{aligned}$$

A monotonically increasing polynomial $x(t)$ satisfying the end-point constraints can be fit to (x_i, t_i), $i = 1, \ldots, n-1$, as discussed in section 2. Similarly we can fit a monotonically decreasing polynomial $y(t)$ that satisfies the end-point constraints. A sample polynomial fit is using our method is shown in figure 2.

4. HISTOGRAM FITTING

Let the normalized histogram data be (y_i, t_i), $i = 0, \ldots, n-1$, where f_i is the normalized frequency at t_i, and $f_0 = f_{n-1} = 0$. The problem is to fit a curve $y(t)$ such that (i) $y(t) \geq 0$, (ii) area of $y(t)$ within the interval $[t_0, t_{n-1}]$ is equal to 1, and (iii) $y(t)$ minimizes the sum of squared residuals.

Since the density function is non-negative, the integral of the density function is a monotonically increasing function. Let (Y_i, t_i), $i = 0, \ldots, n-1$, be the normalized cumulative frequencies, such that $Y_0 = 0$, and $Y_{n-1} = 1$. The

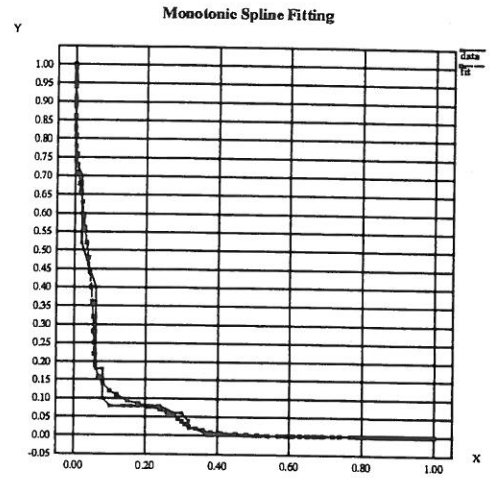

Figure 1: ROC curve fitting with splines. The number of data points, n, was 88 and m was 13. The spline fit required 49 MINOS iterations.

problem is to find a monotonically increasing function $Y(t)$, $t_0 \leq t \leq t_{n-1}$ such that $Y(t_0) = 0$, and $Y(t_{n-1}) = 1$. Again, spline or polynomial fitting can be done using the method discussed in section 2. Finally, since $y(t) = Y'(t)$, we can find the function $y(t)$ that fits the normalized histogram by computing the derivative of function $Y(t)$ that was fitted to the cumulative histogram. Since the cumulative function, $Y(t)$, is non-negative and satisfies the condition $Y(1) = 1$, we are guaranteed that the area under the density function $y(t)$ is equal to 1. The derivatives can be computed using equation 3. In figure 3 we show the result of application of our algorithm on histogram data.

5. DISCUSSION

In this paper we gave algorithms for fitting monotonic splines and polynomials with end-point constraints to ROC data. The monotonicity constraints amounted to constraining the spline coefficients to be in increasing (or decreasing) order and the end-point constraints were incorporated by one linear equality for each endpoint. The constrained regression problem was thus posed as a least-squares problem with linear constraints. We gave detailed algorithms and showed results using our implementation.

To fit a spline to a density function we fit a monotonically increasing spline to the cumulative density function, which is monotonically increasing, and take the derivative of the fitted spline function. The end-point constraints are that the density function value at the left end point (in case of one variable) is 0, and at the right end point is 1. Thus histograms can approximated by smooth splines using the methodology described in this paper.

Currently the order of splines, and the knot locations are provided by the user (we used second-degree basis splines, and uniform knot locations). The optimal order and knot location problem is a model selection problem that can be addressed using a cross validation approach [13], or MDL/AIC method [14, 15], or a Bayesian approach. Furthermore,

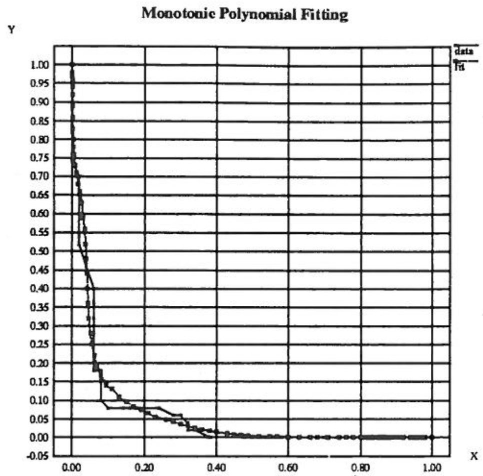

Figure 2: ROC curve fitting with polynomials. The number of data points, n, was 88 and m was 15. The polynomial fit took MINOS 696 iterations.

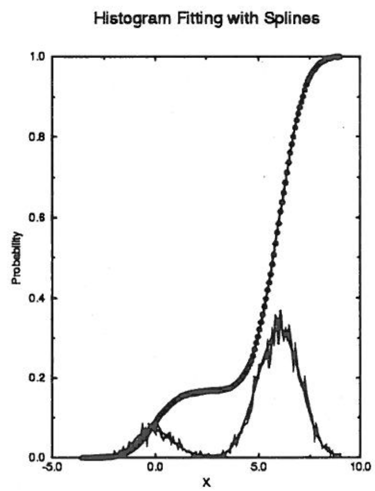

Figure 3: Histogram fitting with splines.

confidence intervals can be estimated using bootstrap techniques.

Acknowledgement: We would like to thank Ken Thornton for discussions on a different formulation of the problem; this work was done while Tapas Kanungo was visiting Bell Labs, Murray Hill, and he would like to thank Henry Baird for inviting him there.

6. REFERENCES

[1] T. Kanungo, M. Y. Jaisimha, J. Palmer, and R. M. Haralick. A quantitative methodology for analyzing the performance of detection algorithms. In *Proc. of IEEE Int. Conf. on Computer Vision*, pages 247–252, Berlin, May 1993.

[2] T. Kanungo, M. Y. Jaisimha, J. Palmer, and R. M. Haralick. A methodology for quantitative performance evaluation of detection algorithms. *IEEE Trans. on Image Processing (to appear)*, December 1994.

[3] T. Kanungo. On optimal operating points. Technical Report ISL-TR-94-07, University of Washington, Seattle, WA, 1994.

[4] J. Friedman and R. Tibshirani. The monotone smoothing of scatterplots. *Technometrics*, 26(3):243–250, 1984.

[5] J. O. Ramsay. Monotone regression splines in action. *Statistical Science*, 3(4):425–461, 1988.

[6] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming.* The Scientific Press, South San Francisco, 1993.

[7] C. de Boor. *A Practical Guide to Splines.* Springer-Verlag, New York, 1978.

[8] P. Dierckx. *Curve and Surface Fitting with Splines.* Clarendon Press, New York, 1993.

[9] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, 17(2):238–246, 1980.

[10] B. A. Murtagh and M. A. Saunders. Minos 5.1 user's guide. Technical Report SOL 83-20R, Stanford University, 1987.

[11] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. User's guide for NPSOL (version 4.0): A fortran package for nonlinear programming. Technical Report SOL 86-2, Stanford University, 1986.

[12] J. J. Moré and S. J. Wright. *Optimization Software Guide.* SIAM, Philadelphia, PA, 1993.

[13] S. N. Wood. Monotonic smoothing splines fitted by cross validation. *SIAM Journal on Scientific Computing*, 15(5):1126–1133, 1994.

[14] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15. World Scientific Series in Computer Science, 1989.

[15] F. S. Cohen and J. Y. Wang. Part i: modeling image curves using invariant 3-d object curve models—a path to 3-d recognition and shape estimation from contours. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 16(1):1–12, 1994.