

Non Parametric Regression

Statistics 135

Autumn 2005



Nonparametric Regression

Reference: Hastie T, Tibshirani R, Friedman J (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction.

In the regression type procedures discussed so far, the form of the regression function has been specified. For example, we might use

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

a cubic polynomial.

However in some situations we don't have enough information to make an assumption like this, or we don't want to. Instead we might want to only assume

$$y = f(x) + \epsilon$$

With some smoothness assumptions on $f(x)$, such as continuity of the regression function its derivatives.

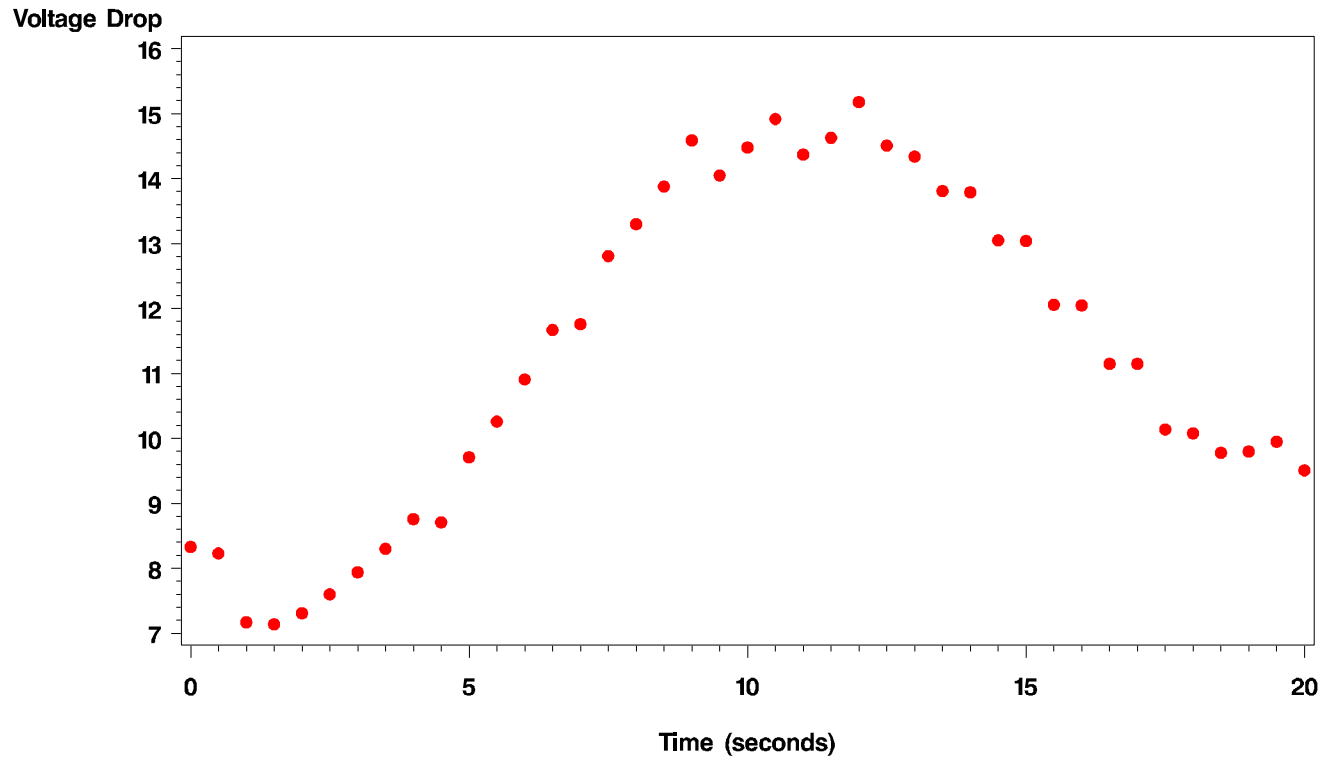
A common assumption that is used is that $f(x)$, $f'(x)$, and $f''(x)$ are all continuous.

So we would like to find an estimate of $f(x)$ based on the data.

Example: Battery Voltage Drop

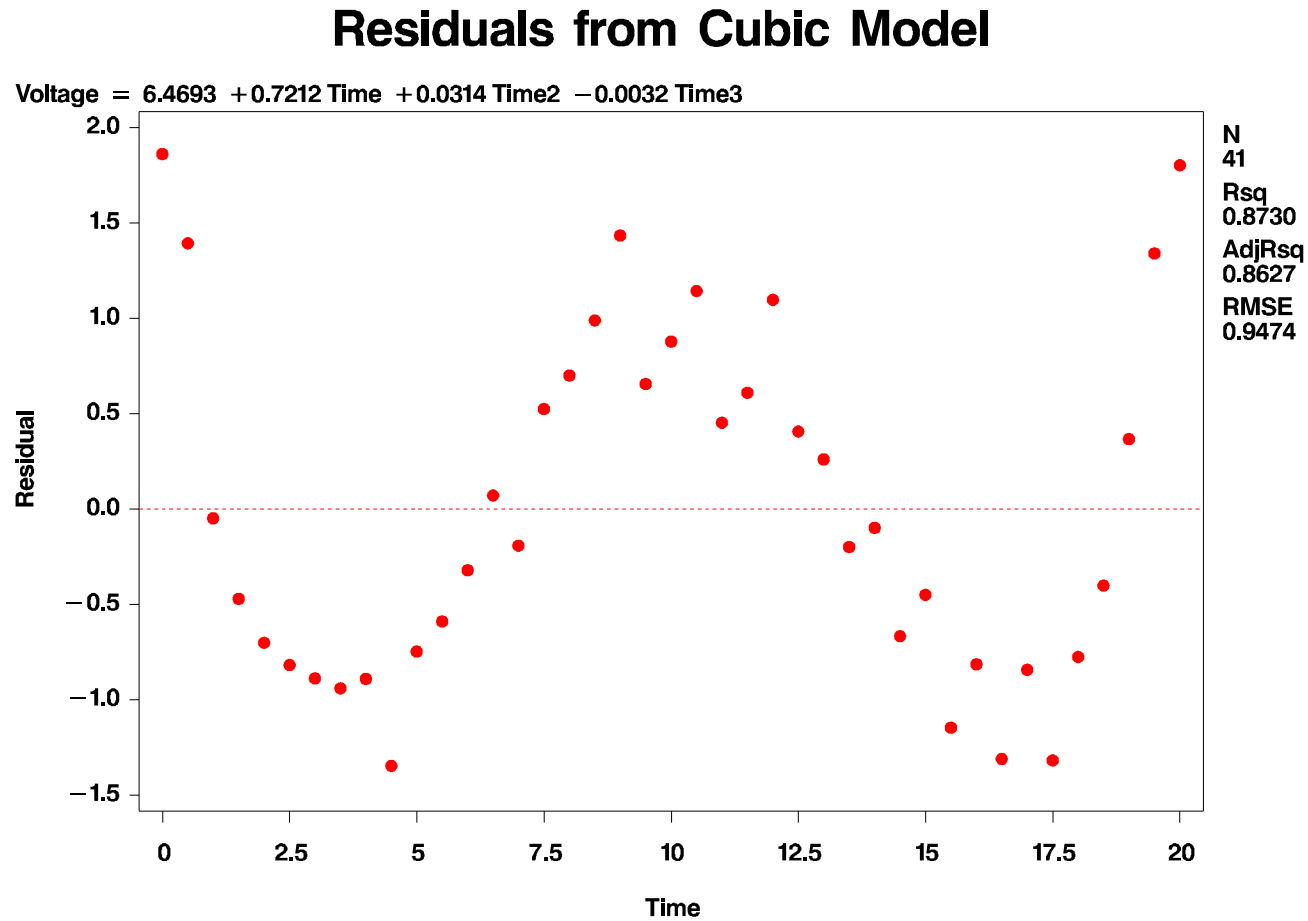
The battery voltage drop in a guided missile motor was observed over the first 20 seconds of the launch. The data was collected to develop a digital-analog simulation model of the missile.

Voltage Drop vs Time

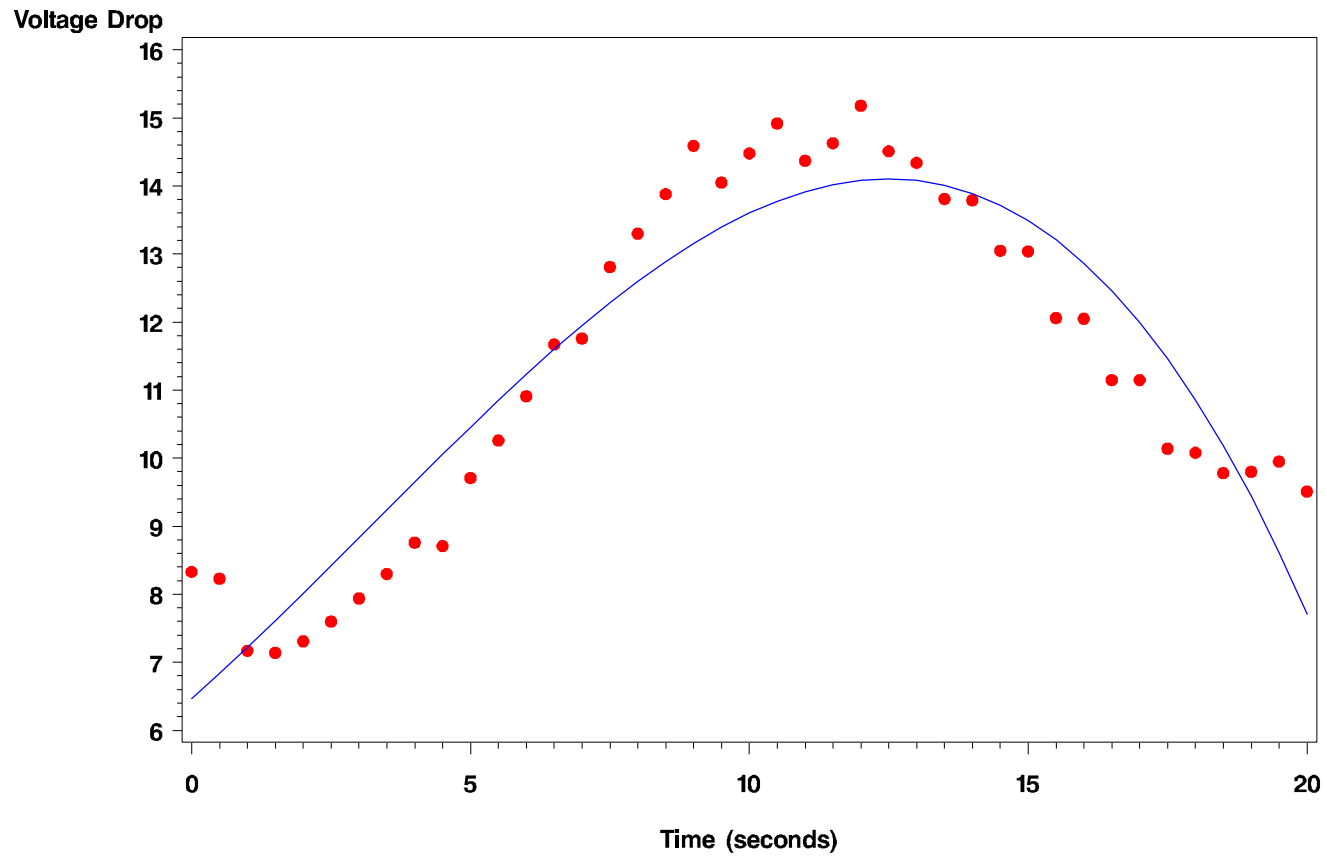


Lets look at a cubic fit of this data.

The plots of the fits and the residuals from this model suggests that this is not a good model



Voltage Drop vs Time – Quartic Fit

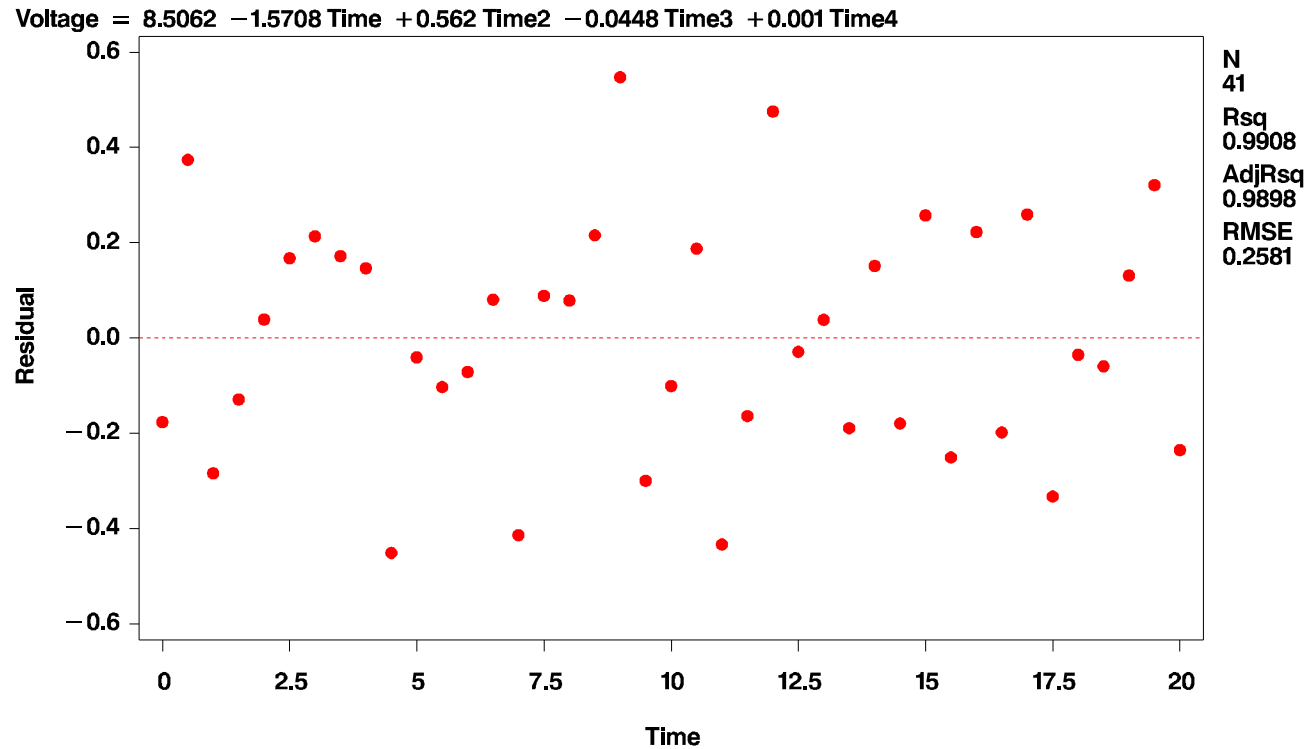


However the quartic model

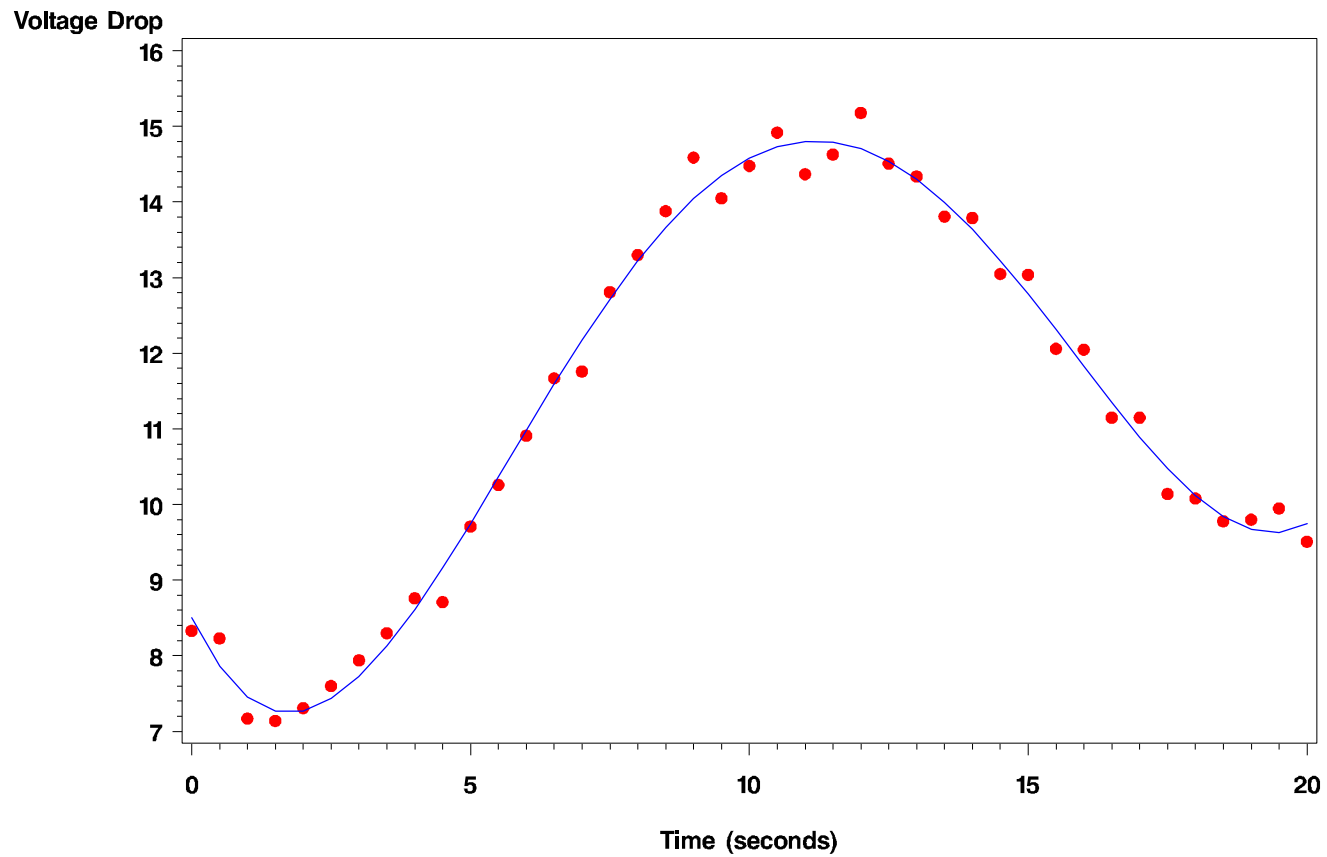
$$y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \epsilon$$

seems reasonable based on the the following two plots.

Residuals from Quartic Model



Voltage Drop vs Time – Quartic Fit



Smoothing Splines

Instead of using the quartic model, let's investigate a cubic smoothing spline. These come considering the following problem: among all functions $f(x)$ with 2 continuous derivatives minimize the penalized residual sum of squares

$$RSS(f, \lambda) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b (f''(t))^2 dt$$

where λ is a fixed *smoothing parameter*. The first term measures the closeness of the data where the second term penalizes the curvature of the function, and λ establishes the tradeoff between the two. The effect ranges from

- $\lambda = 0$: f can be any function that interpolates the data
- $\lambda = \infty$: the least squares straight line fit, since no second derivative can be tolerated.

So the resulting estimates may range from very rough to very smooth.

Even though the criterion is defined on an infinite-dimensional function space (a Sobolov space of functions for which the second term is defined), it ends up it has an explicit, finite-dimensional unique minimizer.

The optimizer is a natural cubic spline with knots at the unique values of $x_i, i = 1, \dots, n$. This is a piecewise cubic polynomial.

Since the solution is a natural spline, we can write it as

$$f(x) = \sum_{j=1}^n N_j(x)\theta_j$$

where the $N_j(x)$ are an n -dimensional set of basis functions.

Thus the criterion thus reduces to

$$RSS(\theta, \lambda) = (\mathbf{y} - \mathbf{N}\theta)^T (\mathbf{y} - \mathbf{N}\theta) + \lambda \theta^T \Omega_N \theta$$

where $\{\mathbf{N}\}_{ij} = N_j(x_i)$ and

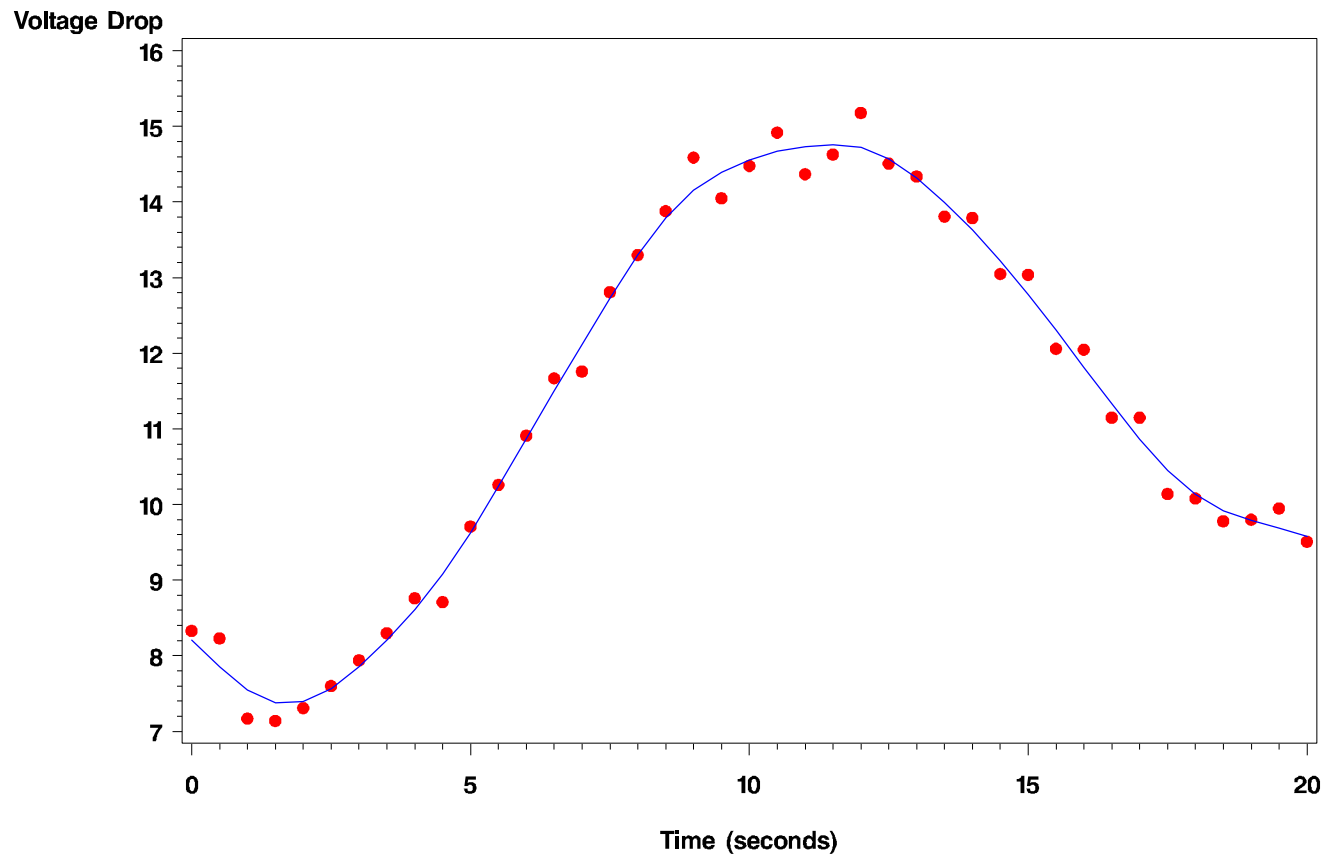
$$\{\Omega_N\}_{jk} = \int N_j''(t) N_k''(t) dt$$

It is possible to show that the solution is

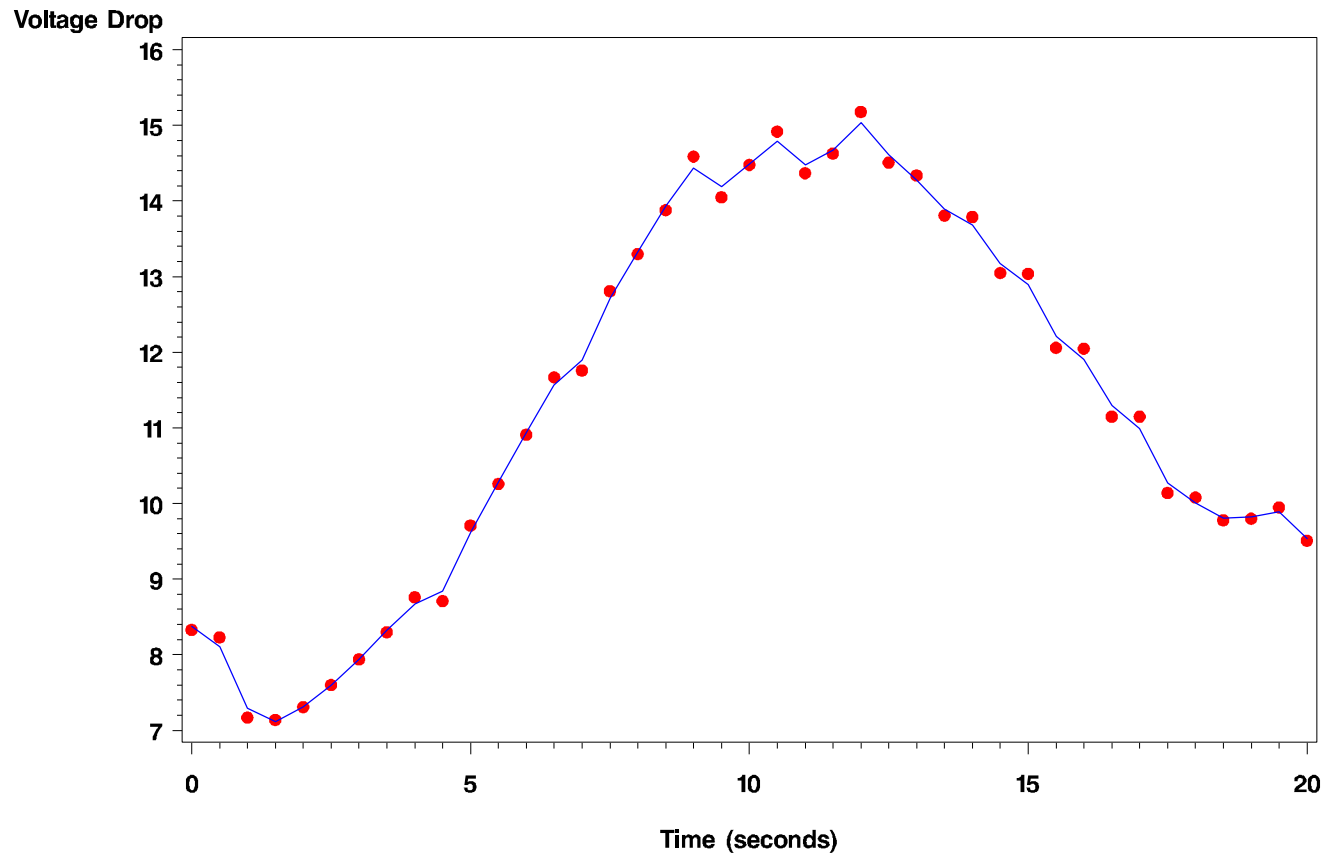
$$\hat{\theta} = (\mathbf{N}^T \mathbf{N} + \lambda \Omega_N)^{-1} \mathbf{N}^T \mathbf{y}$$

To see the effect of the choice of λ , lets look at a couple of fits generated by **SAS**. The first has lambda chosen by Generalized Cross Validation and the second with a fixed lambda.

Voltage Drop vs Time – Spline fit (lambda = 0.89 – GCV)



Voltage Drop vs Time – Spline fit ($\lambda = 0.07$ – $DF=30$)



Generalized Cross Validation is an approach to having the data pick the smoothing parameter. It looks at the bias and variance of prediction and tries to minimize of the two of these. (It has a similar feel to C_p for model selection.)

In **SAS** to specify the amount of smoothness, you don't specify λ but instead a degrees of freedom parameter. This is based on the fact that the fitted values satisfy

$$\begin{aligned}\hat{\mathbf{f}} &= \mathbf{N}(\mathbf{N}^T\mathbf{N} + \lambda\Omega_N)^{-1}\mathbf{N}^T\mathbf{y} \\ &= \mathbf{S}_\lambda\mathbf{y}\end{aligned}$$

i.e. the fits are a linear combination of the y s

By analogy with linear regression, the degrees of freedom for a smoothing spline can be defined by

$$df_\lambda = \text{trace}(\mathbf{S}_\lambda)$$

This idea can be extended to other nonparametric regression procedures which is why **SAS** does it this way.

The relationship between λ and df_λ is problem specific but only depends on the x s.

A **SAS** approach for fitting is with PROC GAM, the procedure for fitting Generalized Additive Models.

The code for fitting

```
PROC GAM DATA = missile2 ;  
  MODEL Voltage = SPLINE(Time) / METHOD = GCV ;  
  OUTPUT OUT = spline PREDICTED RESIDUAL;
```

```
PROC GAM DATA = missile2 ;  
  MODEL Voltage = SPLINE(Time, DF=30) ;  
  OUTPUT OUT = spline2 PREDICTED RESIDUAL;
```

If DF is not defined and METHOD = GCV is not used, DF=4 is used. The OUTPUT options are included to generate the plots. They are not needed to run the procedure. The output from GAM is

The GAM Procedure

Dependent Variable: Voltage

Smoothing Model Component(s): spline(Time)

Summary of Input Data Set

Number of Observations	41
Number of Missing Observations	0
Distribution	Gaussian
Link Function	Identity

Iteration Summary and Fit Statistics

Final Number of Backfitting Iterations	2
Final Backfitting Criterion	0
The Deviance of the Final Estimate	2.0316861891

The local score algorithm converged.

Regression Model Analysis
Parameter Estimates

Parameter	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	9.51111	0.08013	118.69	<.0001
Linear(Time)	0.18113	0.00690	26.26	<.0001

Smoothing Model Analysis
Fit Summary for Smoothing Components

Component	Smoothing Parameter	DF	GCV	Num Unique Obs
Spline(Time)	0.887627	9.233560	0.094013	41

Smoothing Model Analysis
Analysis of Deviance

Source	DF	Sum of Squares	Chi-Square	Pr > ChiSq
Spline(Time)	9.23356	212.495030	3113.2861	<.0001

Regression Model Analysis
Parameter Estimates

Parameter	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	9.51111	0.06099	155.95	<.0001
Linear(Time)	0.18113	0.00525	34.51	<.0001

Smoothing Model Analysis
Fit Summary for Smoothing Components

Component	Smoothing Parameter	DF	GCV	Num Unique Obs
Spline(Time)	0.070124	29.000000	0.162103	41

Smoothing Model Analysis
Analysis of Deviance

Source	DF	Sum of Squares	Chi-Square	Pr > ChiSq
Spline(Time)	29.00000	214.131342	5415.9193	<.0001

Note that **SAS** removes the linear trend before fitting the spline which takes one degree of freedom. This is why the second example has 29, not 30 degrees of freedom.

PROC GAM will do many analyses. In addition to using smoothing splines, it will also do loess fits as well.

This can be done by

```
PROC GAM DATA = missile2 ;  
  MODEL Voltage = LOESS(Time) / METHOD = GCV ;  
  OUTPUT OUT = loess PREDICTED RESIDUAL;
```

If you wish to set the smoothness via DF, it works the same way as with SPLINE.

Regression Model Analysis
Parameter Estimates

Parameter	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	11.32244	0.03984	284.18	<.0001

Smoothing Model Analysis
Fit Summary for Smoothing Components

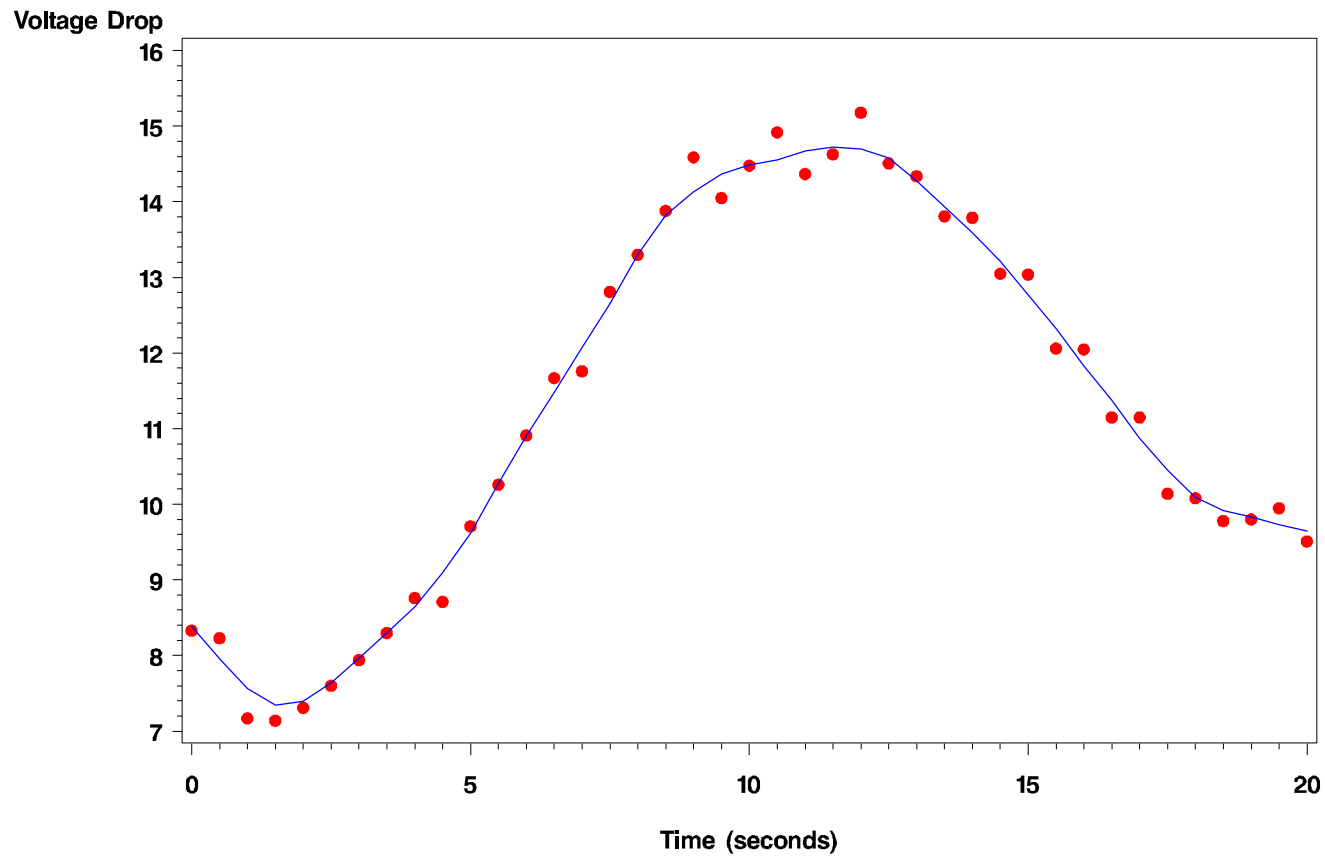
Component	Smoothing Parameter	DF	GCV	Num Unique Obs
Loess(Time)	0.158537	9.566316	0.002465	41

Smoothing Model Analysis
Analysis of Deviance

Source	DF	Sum of Squares	Chi-Square	Pr > ChiSq
Loess(Time)	9.56632	259.626796	3989.0744	<.0001

The fits in this case are similar to the spline fits

Voltage Drop vs Time – Loess fit (GCV)



Note that this procedure is for Generalized Additive Models. The Generalized means the same thing as in Generalized Linear Models. It is possible to assume that the conditional distribution of the response variable given the predictors doesn't have to be normal. The option `DIST` of the `MODEL` statement allows for `GAUSSIAN` (default), `BINARY`, `BINOMIAL`, `GAMMA`, `IGAUSSIAN`, and `POISSON` distributions to be used. In all cases, the canonical link is used. Not being able to use non canonical links in this case should not be a problem.