Local weighted least squares: Quadratic loess smoother

An example of weighted least squares fitting of data to a simple model for the purposes of simultaneous smoothing and interpolation is the quadratic loess smoother.

In one dimension, this can be used to smooth, filter or interpolate (possibly all at once) a time series of values that may or may not be at regular sampling intervals.

An application in two or more dimensions could be to produce a gridded analysis or climatology from a set of data observed at irregularly spaced locations and times, such as a set of shipboard hydrographic observations of temperature, salinity, or other ocean properties.

The “model” as originally described is a local quadratic function, which can be written in terms of coordinates centered at the estimation point, \( x_o, y_o \)

\[
y = a_1 + a_2 (x - x_o) + a_3 (y - y_o) + a_4 (x - x_o)(y - y_o) + a_5 (x - x_o)^2 + a_6 (y - y_o)^2
\]

which has the design matrix

\[
E = \begin{bmatrix}
1 & x_i - x_o & y_i - y_o & (x_i - x_o)(y_i - y_o) & (x_i - x_o)^2 & (y_i - y_o)^2 \\
\end{bmatrix}
\]

with coefficients \( \mathbf{a} = [a_1 : : a_6] \) and data \( \mathbf{d} = [d_1 : : d_N] \)

The flexibility to choose the coefficients of any linear model (in the coefficients) means the least squares fitting approach can be tailored to meet a user’s notion of what constitutes a rational model based on some a priori knowledge of the processes being observed and modeled.

An example of this is the Climatology of the Australian Regional Seas (CARS):
Additional terms in the “model” can be included provided they have a form that can be expressed with the design matrix, coefficients $a_k$, and data coordinates (including, e.g. observation times $t_i$).

Length and time scales of variability

The quadratic loess smoother requires an a priori choice be made for the scales (usually length or time) to apply in the selection of the smoothing weights.

The smoother can be interpreted as a filter, since the linear weighting procedure is effectively implemented as a convolution of the weights with the data. It is more general in the sense that the data do not have to be at regular intervals, because the weights are computed simply as a function of the normalized distance function $r$.

It has been shown, empirically, that the effective cutoff frequency $f_c$ of the quadratic loess smoother, when it is interpreted as a filter, is $f_c \approx L^{-1}$ where $L$ is the half width (i.e. the normalization scale) used in the loess smoother.

If the loess smoother is to be used to deliberately remove certain scales of variability (i.e. as a filter) then selection of $L$ is straightforward.

However, if the objective is to use the loess smoother to do the best possible job of interpolating gaps in the data, then the smoothing scale should be adapted to the natural length or time scales of variability in the underlying physical process being observed.

The weighting is not associated with errors in the data but the recognition that data some distance away from the estimation location are less representative of the state because the field that is being estimated is varying in space or time.

Weighting

The weighting is applied to the model equations from which we seek the least squares best solution for the parameters $a$.

The rows of the matrix equation $Ea = d$ are weighted by $w_i$ that can be summarized as a weighting vector $w$

$$\text{diag}(w)Ea = \text{diag}(w)d$$
$$WEa = Wd$$
$$\hat{E}a = Wd$$

which can be solved (in the least squares sense) with

$$>> a=Ehat\backslash(W*d);$$

The classic quadratic loess smoother uses the weighting function:
\[ w = (1 - r^3)^3 \quad \text{in} \quad r < 1 \]

where \( r \) might be normalized one of two ways:

Method (1) uses a fixed smoothing scale.

Method (2) adapts to the density of the available data.

(1) \( r \) is a normalized Cartesian distance with prescribed smoothing scale \( L \)

\[
\left( \frac{x - x_o}{L} \right)^2 + \left( \frac{y - y_o}{L} \right)^2 \right)^{1/2}
\]

(2) \( r \) is normalized differently for each estimation point \( x_o, y_o \) after finding the distance \( r_{max} \) that encloses the nearest \( N \) data points

\[
\begin{align*}
\begin{aligned}
r^* &= \left( (x - x_o)^2 + (y - y_o)^2 \right) \\
r &= \text{sort}(r^*) \\
r_{max} &= r^*(N) \\
r &= \frac{r^*}{r_{max}}
\end{aligned}
\end{align*}
\]

Since the data coordinates are transformed to be with respect to the estimation location, \( x_o, y_o \), the final loess estimate is simply \( a_L \).

Two-dimensional spatial mapping using a loess filter is demonstrated in Matlab script \texttt{jw_loess2d.m}.