1 Nonparametric Regression (Cont'd)

1.1 Lowess Curves

- Another **nonparametric regression** method is the ***lowess*** (or ***loess***) curve, or **"locally weighted scatterplot smoother."**

- As before, our goal is to fit a model of the form

$$Y_i \;=\; f(X_i) \,+\, \epsilon_i\,,$$

  without specifying a form for $f(x)$.

- The blue curve in Fig. 1 is a **lowess** curve for Fairbanks, Alaska annual mean temperatures for the years 1930 - 2008.

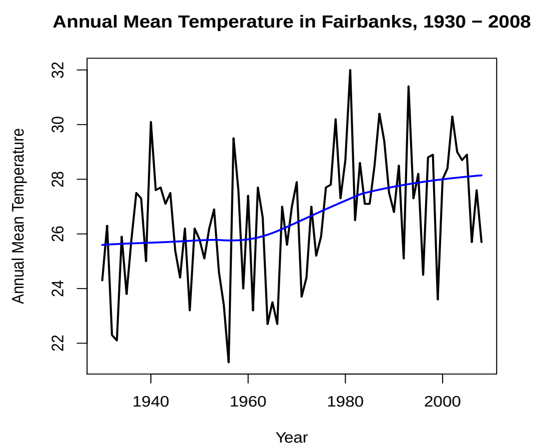

Figure 1

- We'll see how to obtain the value of the **lowess** curve $\boldsymbol{\hat{f}(x)}$ at a *single* point $\boldsymbol{x = X_h}$.

  In practice, the value of $\boldsymbol{\hat{f}(x)}$ would need to be obtained at *numerous* points in the manner described, and these values connected to produce a curve like in Fig. 1.

- The ***lowess*** curve has two "parameters":

  1. A ***neighborhood size q***, expressed as a proportion of the full data set, indicating the proportion of points from the full data set that influence the curve at X_h.

2. A ***weight function**** that determines a ***weight*** $\boldsymbol{w_i}$ for each $Y_1, Y_2, \ldots, Y_n$ to be used to determine $\hat{f}(x)$ at $x = X_h$.

   The **weights** $\boldsymbol{w_i}$ decrease with the distance of X_i away from X_h (but don't necessarily sum to 1).

   Only the **closest** $\boldsymbol{q}$ **percent** of observations to $\boldsymbol{X_h}$ have nonzero weights. This is the "**lo**cal" part of "**lo**wess".

   Observations **farther** from $\boldsymbol{X_h}$ have **smaller weights**. This is the "**we**ighted" part of "lo**we**ss".

   *Usually the so-called *tricube weight function* is used:

$$w_i = \begin{cases} \left(1 - \left(\frac{d_i}{d_q}\right)^3\right)^3 & \text{if } d_i < d_q \\ 0 & \text{if } d_i \geq d_q \end{cases}$$

   where d_i is the distance from X_h to X_i and d_q is the distance from X_h to the farthest point from X_h that's among the closest q percent of points.

- The value of the ***lowess*** curve $\boldsymbol{\hat{f}(x)}$ at $\boldsymbol{x = X_h}$ is the *fitted value* $\hat{Y}_h$ at X_h from fitting a **simple linear** (or **quadratic polynomial**) **regression model** using so-called ***weighted least squares***.**

**The *weighted least squares estimates* b_0 and b_1 of the unknown parameters β_0 and β_1 are the values that minimize

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i (Y_i - \beta_0 - \beta_1 X_i)^2$$

as a function of β_0 and β_1.

- The **window size** $\boldsymbol{q}$ is a ***tuning parameter*** that controls the degree of "wiggle" in the *lowess* curve (i.e. controls the ***bias-variance trade-off***).

  A **smaller window** results in a **more "wiggly"** (**high variance** but **low "bias"**) *lowess* curve that **fits** the data **better** but **risks overfitting**. See Fig. 2.

  A **larger window** results in a **"smoother"** (**low variance** but **high "bias"**) curve but with potentially greater **"bias"** (analogous to model *misspecification*). See Fig. 2.

  The optimal value of $\boldsymbol{q}$ can be selected via **cross-validation**.

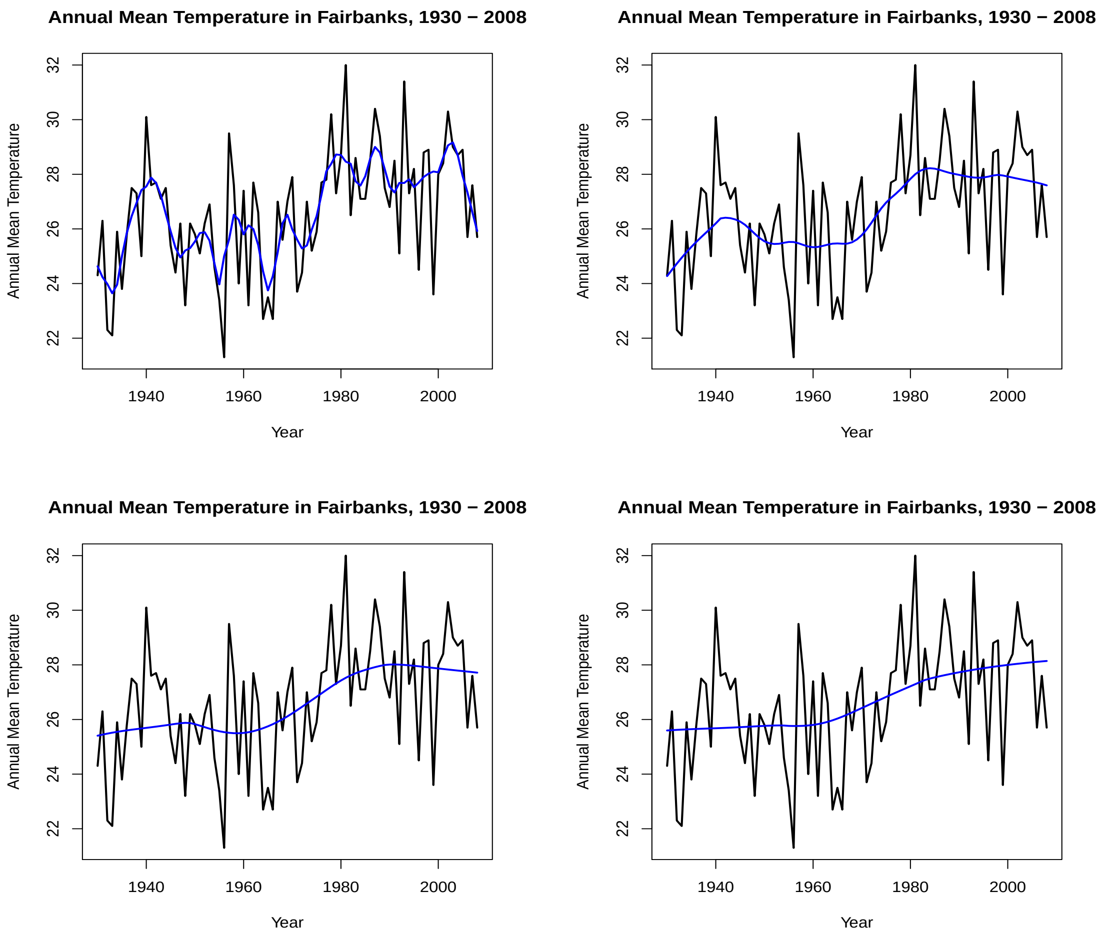

Figure 2: Lowess curves for annual mean temperatures in Fairbanks, Alaska over the years 1930 - 2008, with neighborhood sizes $q = 0.1$ (upper left), $q = 0.3$ (upper right), $q = 0.5$ (lower left), and $q = 0.67$ (lower right).

- A **moving average** (Class Notes 25) can be viewed as a *special case* of a **lowess** curve. See the next example.

> **Example 1.1** Recall, if we fit a **regression model** with **just** an **intercept**,
>
> $$Y_i \;=\; \beta_0 + \epsilon_i \,,$$

the estimate b_0 of β_0 minimizes

$$Q(\beta_0) \;=\; \sum_{i=1}^{n} (Y_i - \beta_0)^2$$

as a function of β_0. Setting the derivative $dQ/d\beta_0$ equal to zero and solving gives

$$b_0 \;=\; \bar{Y}\,,$$

so the **fitted model** is

$$\hat{Y} \;=\; b_0 \;=\; \bar{Y}\,,$$

and the **fitted value**, for every $\boldsymbol{X_h}$, is $\boldsymbol{\hat{Y}_h = \bar{Y}}$.

Now consider *time series* data. If for each time point $\boldsymbol{X_h}$, the **intercept-only model** is fitted to observations within a period of time $\boldsymbol{q}$ **before** *and* **after** time $\boldsymbol{X_h}$, the resulting *fitted values* $\hat{Y}_h$ form an (intercept-only) **lowess** curve that's also a **simple moving average**.***

***More precisely, this **lowess** curve uses **weights**

$$w_i \;=\; \begin{cases} 1 & \text{if } d_i \le q \\ 0 & \text{if } d_i > q \end{cases}$$

where d_i is the number of time points from X_h to X_i, and it can be shown (using *weighted least squares*) that the **lowess** curve $\boldsymbol{\hat{f}(x)}$ at $\boldsymbol{x = X_h}$ is the average

$$\hat{f}(X_h) \;=\; \frac{1}{2q+1} \sum_{i=h-q}^{h+q} Y_i\,.$$