Andrew Gard - equitable.equations@gmail.com

Fitting a loess curve

Fitting a loess curve

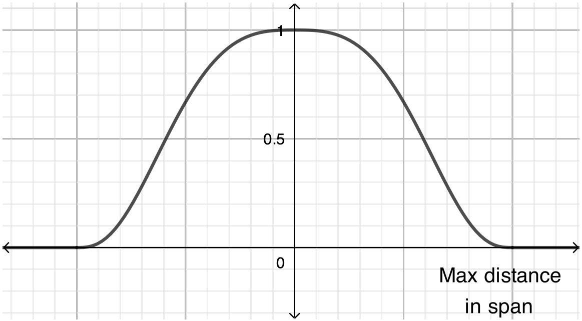
Given n data points $(x_1, y_1), \cdots, (x_n, y_n)$, a hyperparameter α (denoted in R by `span`) with $0 < \alpha \leq 1$, and a target explanatory variable value x:

1. Consider the fraction α of points in the data set that are nearest to x.
2. Weight points close to x more than points farther away.
3. Fit a quadratic model using that weighted set.
4. Use that model to predict the value of the response variable for the explanatory value x.

This process is known as loess regression (<u>lo</u>cally <u>e</u>stimated <u>s</u>catterplot <u>s</u>moothing).

For a given value x, the **weighting function** $w(x, x_j)$ for the j^{th} point in the set S of sample points is given by the tricubic polynomial,

$$w(x, x_j) = \left(1 - \left|\frac{x - x_j}{\max_{x_i \in S} |x_i - x_j|}\right|^3\right)^3$$



More fun facts:

- Span values $\alpha > 1$ are also permisable. This has the effect of stretching the tricubic weighting function horizontally so that further points receive more equal weight.

More fun facts:

- Span values $\alpha > 1$ are also permisable. This has the effect of stretching the tricubic weighting function horizontally so that further points receive more equal weight. As $\alpha \to \infty$, all points are used approximately equally and the loess model approaches a simple quadratic.

More fun facts:

- Span values $\alpha > 1$ are also permisable. This has the effect of stretching the tricubic weighting function horizontally so that further points receive more equal weight. As $\alpha \to \infty$, all points are used approximately equally and the loess model approaches a simple quadratic.
- The degree of the local model can be adjusted, though this is seldom desirable.

Loess models are very flexible They're non-parametric and can fit any distribution of data. There are disadvantages as well, however.

Loess models are very flexible They're non-parametric and can fit any distribution of data. There are disadvantages as well, however.

- They aren't transparent or easy to interpret. This is particularly true when multiple explanatory variables are present.

Loess models are very flexible They're non-parametric and can fit any distribution of data. There are disadvantages as well, however.

- They aren't transparent or easy to interpret. This is particularly true when multiple explanatory variables are present.
- They are prone to overfitting.

Loess models are very flexible They're non-parametric and can fit any distribution of data. There are disadvantages as well, however.

- They aren't transparent or easy to interpret. This is particularly true when multiple explanatory variables are present.
- They are prone to overfitting.
- The hyperparameter α must be tuned, which requires knowledge and care.

Loess models are very flexible They're non-parametric and can fit any distribution of data. There are disadvantages as well, however.

- They aren't transparent or easy to interpret. This is particularly true when multiple explanatory variables are present.
- They are prone to overfitting.
- The hyperparameter α must be tuned, which requires knowledge and care.
- They are computationally expensive. Literature typically recommends against its use when $n > 1000$.

.