A regression model describes a relation between a vector of d real valued input variables (features) $x = (x_1, x_2, ..., x_d)$ and a single real valued output variable y. Using a finite number n of training observations (data cases or data points) $(x_{(i)}, y_{(i)})$, $i = 1,2,...,n$ one wants to build a model F that allows predicting the output value for yet unseen input values as closely as possible.

Locally Weighted Polynomials

Locally Weighted Polynomial (LWP) approximation (also called Locally Weighted Regression or Moving Least Squares) (Cleveland & Devlin 1988) is designed to address situations in which the models of global behaviour do not perform well or cannot be effectively applied without undue effort. The LWP approximation is carried out by pointwise fitting of low-degree polynomials to localized subsets of the data. The advantage of this method is that the analyst is not required to specify a global function of the data. However, the method requires considerably higher computational resources for finding the best values of the hyperparameters as well as when LWP is finally used for prediction.

The assumption of the LWP approximation is that near the query point the value of the actual response changes smoothly and can be approximated using a low-degree polynomial. The coefficients of the polynomial are calculated using the weighted least squares method giving the largest weights to the nearest (usually according to the Euclidean distance) data points and the lowest or zero weights to the farthest data points.

Given a local model, e.g. a first degree polynomial,

$$F(x) = a_0 + \sum_{i=1}^{d} a_i x_i \ ,$$

the coefficients a are calculated, minimizing

$$a = \arg\min_a \sum_{i=1}^{n} w(x_{query}, x_{(i)})\left(F(x_{(i)}) - y_{(i)}\right)^2$$

where w is a weight function, x_{query} is the query point nearest neighbours of which will get the highest weights.

The weight function w depends on the Euclidean distance (in scaled space) between the point of interest x_{query} and the points of observations x. One of the most widely used weight functions is the Gaussian weight function. It is defined as follows:

$$w(x_{query}, x_{(i)}) = \exp\left(-\alpha \mu_{(i)}^2\right)$$

where α is a coefficient and the $\mu_{(i)}$ is a scaled distance from the query point to the ith point in the training data set:

$$\mu_{(i)} = \frac{\left\| x_{query} - x_{(i)} \right\|}{\left\| x_{query} - x_{farthest} \right\|}$$

where $\|\cdot\|$ is the Euclidean norm; $x_{farthest}$ is the farthest training point from the point x_{query}. The locality of the approximation is controlled by varying the value of the coefficient α. If α is equal to zero then local approximation transforms into global approximation.

The “best” value of α can also be automatically found using Leave-One-Out Cross-Validation.

References

1. Cleveland W. S., Devlin S. J. Locally weighted regression: An approach to regression analysis by local fitting, Journal of the American Statistical Association, Vol. 83, 1988, pp. 596-610
2. Kalnins K., Ozolins O., Jekabsons G. Metamodels in design of GFRP composite stiffened deck structure. Proceedings of 7th ASMO-UK/ISSMO International Conference on Engineering Design Optimization, Association for Structural and Multidisciplinary Optimization in the UK (ASMO-UK), ISBN: 978-0853162728, Bath, UK, 2008, pp. 263-273