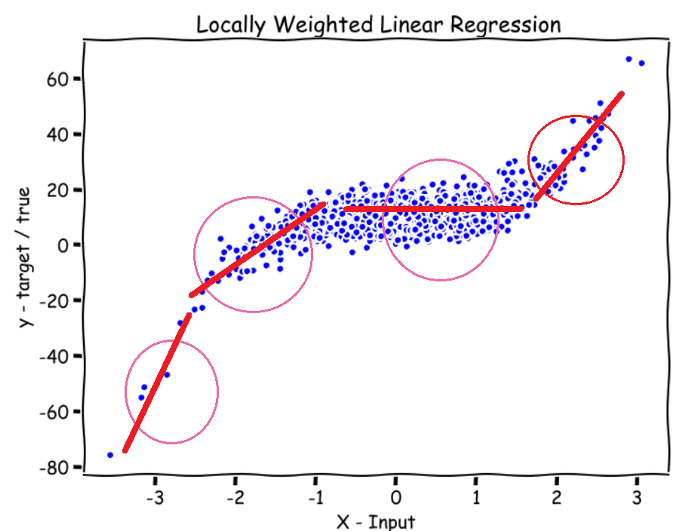
The goal of this regression analysis using tips dataset is to predict the tip amount based on the total bill of the person paying the bill.

**1. Non parametric locally weighted regression.**

Imagine you have a bunch of data points on a graph, and you want to draw a line through them to show how they are related. You might not know exactly what the line should look like, but you think it might be wavy or curvy, rather than a straight line. This is what we call a non-linear relationship between the data points.

One way to draw a line through the data points is to use something called non-parametric locally weighted regression. This is a special way of fitting a line to the data points that takes into account how close each point is to the line.



Imagine you have a special tool that lets you draw a line through the data points, but the line is more influenced by the points that are close to it and less influenced by the points that are farther away. This tool is called a kernel function, and it's kind of like a special pen that gets more or less powerful depending on how close it is to a data point.

When you use non-parametric locally weighted regression, you start by picking a data point on the graph, and then you use the kernel function to draw a line through the data points that is more influenced by the points that are close to the one you picked. You can then move to a different data point and draw a new line through the data points that is more influenced by the points that are close to that one. You can keep doing this for all of the data points on the graph, and eventually you will have a bunch of lines that you can use to fit a curve to the data.

Non-parametric locally weighted regression is a great way to show how data points are related when you're not sure what the relationship looks like. It can help you understand how different variables are connected, and it can even help you make predictions about what might happen in the future.

**2. More details about the Non parametric locally weighted regression.**

Non-parametric locally weighted regression is a method for fitting a regression model to data when the true underlying relationship between the input and output variables is not known a priori. It is a non-parametric method, meaning that it does not make any assumptions about the form of the underlying relationship between the input and output variables.

In locally weighted regression, the model is fit to the data by weighting the points in the data according to their proximity to a given test point. The weight of each point is determined by a kernel function, which is typically a Gaussian function. The model is then fit to the data by minimizing the squared error between the predicted output and the true output, weighted by the kernel function. This results in a model that is more heavily influenced by points that are close to the test point and less influenced by points that are farther away.

The advantage of locally weighted regression is that it can model complex, non-linear relationships between the input and output variables without making any assumptions about the form of the relationship. The disadvantage is that the model may be more sensitive to noise in the data, and may not generalize well to new data.

**3. About Regression**

Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The goal of regression is to predict the value of the dependent variable based on the values of the independent variables.

There are two main types of regression: linear regression and non-linear regression. Linear regression is used to model linear relationships between the dependent and independent variables, where the predicted value of the dependent variable is a linear combination of the independent variables. Non-linear regression is used to model non-linear relationships between the dependent and independent variables.

Regression is used in a wide variety of fields, including economics, finance, psychology, and biology, to name a few. It is a useful tool for understanding how different variables are related, and for making predictions about future outcomes.

**4. why kernel function is needed in Non-parametric locally weighted regression**

In non-parametric locally weighted regression, the kernel function is used to weight the observations in the training data set according to their distance from the input point for which we are trying to make a prediction. The kernel function assigns higher weights to observations that are closer to the input point and lower weights to observations that are further away. This weighting is used to give more emphasis to nearby observations and less emphasis to distant observations when fitting the model.

Kernel smoothing is a technique that can be used in conjunction with locally weighted linear regression to smooth the fit of the model. Kernel smoothing is a useful technique for improving the smoothness and stability of the fit in a locally weighted linear regression model. It can be particularly useful when the training data contains noise or when the model needs to make predictions over a large range of input values.

**5. What is the "local weight" for an observation in Non-parametric locally weighted regression**

the "local weight" for an observation is the weight that is assigned to that observation by the kernel function. The local weight is used to give more emphasis to certain observations and less emphasis to others when fitting the model.The local weights can be thought of as a measure of the influence that each observation has on the model fit at the input point. Observations with higher local weights will have more influence on the model fit, while observations with lower local weights will have less influence. This allows the locally weighted regression model to adapt to the local structure of the data, rather than relying on a global fit that may not be appropriate for all regions of the input space.

**6. What are model parameters in locally weighted linear regression**

the model parameters are the coefficients of the linear regression model that is fit to the training data at each iteration. These coefficients determine the slope and intercept of the line that best fits the training data.

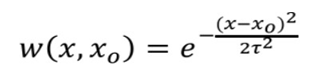
**7.Algorithm**

1. Read the Given data Sample to X and the curve (linear or non linear) to Y

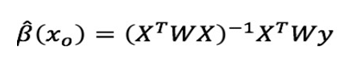
2. Set the value for Smoothening parameter or Free parameter say τ

3. Set the bias /Point of interest set x0 which is a subset of X

4. Determine the weight matrix using the Kernel function :



5. Determine the value of model term parameter β using:



6. Prediction = x0\*β

**8. Explanation of each line of code:**

The first 4 lines import the necessary libraries for the script:

matplotlib.pyplot is a library for creating plots and charts in Python.

pandas is a library for data manipulation and analysis.

numpy is a library for numerical computing in Python.

The kernel function is defined. It takes 3 arguments: a point (point), a matrix of input data (xmat), and a kernel parameter (k). The function calculates a weight matrix (weights) of size m x m, where m is the number of rows in xmat. The weight of each point in xmat is calculated using the Gaussian kernel formula, which is defined as exp(-(point - xmat[j])^2 / (2\*k^2)), where j is the index of the point in xmat.

The localWeight function is defined. It takes 4 arguments: a point (point), a matrix of input data (xmat), a matrix of output data (ymat), and a kernel parameter (k). The function calculates the locally weighted linear regression coefficients for the given point using the input and output data and the kernel parameter. The function first calls the kernel function to get the weight matrix for the given point, and then uses this weight matrix to calculate the locally weighted linear regression coefficients using the weighted least squares method.

The localWeightRegression function is defined. It takes 3 arguments: a matrix of input data (xmat), a matrix of output data (ymat), and a kernel parameter (k). The function calculates the locally weighted linear regression predictions for all points in xmat using the input and output data and the kernel parameter. The function does this by calling the localWeight function for each point in xmat.

The tips.csv file is read into a Pandas DataFrame using the read\_csv function. The total\_bill and tip columns are extracted from the DataFrame and converted into numpy arrays.

The mbill and mtip matrices are created from the bill and tip arrays, respectively. The one matrix is created as a matrix of all ones with the same number of columns as mbill. The X matrix is created by stacking one and mbill horizontally using the hstack function.

The localWeightRegression function is called to get the locally weighted linear regression predictions for all points in X. The kernel parameter is set to 0.5.

The SortIndex array is created by sorting the second column of X in ascending order. The xsort array is created by sorting X using the indices in SortIndex.

A figure and subplot are created using matplotlib.pyplot, and the scatter plot of the input data (bill and tip) is plotted on the subplot. The locally weighted linear regression predictions are plotted on the same subplot using the plot function. The x-axis is labeled "Total bill" and the y-axis is labeled "Tip".

The plot is displayed using the show function.

**9. About Tips Data Set.**

The Tips dataset is a small dataset included with the seaborn library in Python that is used for demonstrating regression plots. It contains 244 observations of 2 numeric variables, total bill and tip, and 1 categorical variable, gender. The total bill amount is the amount of money spent at a restaurant, and the tip amount is the additional money given as a tip to the waiter or waitress.

*Description of all of the columns in the Tips dataset:*

**total\_bill:** The total amount of money spent at the restaurant, in US dollars.

**tip:** The amount of money given as a tip, in US dollars.

**gender:** The gender of the person who paid the bill. This can be either "Male" or "Female".

**smoker:** Whether or not the person who paid the bill was a smoker. This can be either "Yes" or "No".

**day:** The day of the week on which the bill was paid. This can be one of "Thur", "Fri", "Sat", or "Sun".

**time:** The time of day at which the bill was paid. This can be either "Lunch" or "Dinner".

**size:** The size of the party that dined at the restaurant. This is an integer representing the number of people in the party.

**Note:** of the 7 columns, we are only using the first 2 columns that is total\_bill and tip. The goal of a regression analysis using this dataset is to predict the tip amount based on the total bill of the person paying the bill.