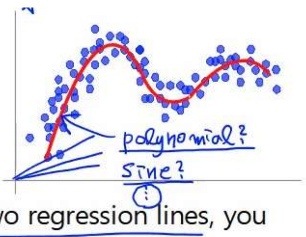
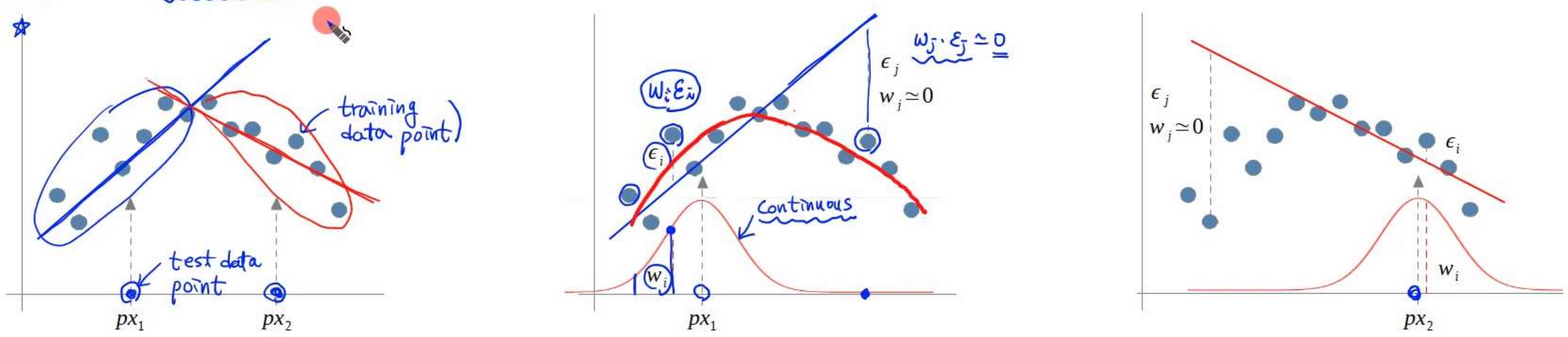
Locally Weighted Regression

Tuesday 21 May 2024 7:44 AM

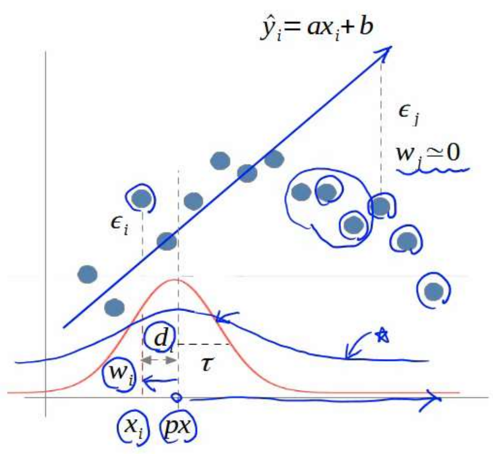
- To learn the data in the figure on the right, nonlinear regression is needed. We need to choose a nonlinear function and fit that function to the data. Polynomials, sine, log, etc. can be chosen as a nonlinear function. However, it is not easy to choose an appropriate nonlinear function in advance. This method is called a parametric method.
- LWR is a type of non-parametric method that performs nonlinear regression without assuming a nonlinear function.



- To learn the data in the figure below, nonlinear regression is needed. If you want to perform linear regression using two regression lines, you can fit the blue line to the data in the left region and the red line to the data in the right region.
- To predict the test data point px1, calculate the weights w of each data point using a normal distribution with mean px1. Then, linear regression is performed by weighting the errors. The data points on the left have high weight, but the data points on the right have low weight. Therefore, linear regression is performed mainly using the data on the left, and a blue straight line is generated. We can use this line to predict the px1. Conversely, test data point px2 can be predicted using the red straight line.
- LWR is a non-parametric method capable of nonlinear regression, but it takes a long time because it must be performed for each test data point. LWR is a lazy learner like KNN.



- Calculate the distance d between the test data point px and all training data points, and calculate weight w for each data point with a normal distribution for d.
- The standard deviation, tau of the normal distribution can be used to adjust the range of neighbors. The tau is a hyper-parameter.
- Applying weights to the OLS cost function produces the weighted cost function used in Locally Weighted Regression.
- The weighted cost function can be written in two forms as follows.



- Weighted cost function

• weight function

$$d_i = |px - x_i|$$

$$w_i = \exp\left(-\frac{d^2}{2\tau^2}\right) \qquad \begin{cases} d_i \to 0 : w_i \to 1 \\ d_i \to \infty : w_i \to 0 \end{cases}$$

• Cost function

weights

$$\min_{w,b} \sum_i w_i \epsilon_i^2 = \min_{w,b} \sum_i w_i (y_i - \hat{y}_i)^2$$

$$= \min_{w,b} \sum_i w_i (y_i - a x_i - b)^2$$

$$w_i (y_i - a x_i - b)^2 = (\overbrace{\sqrt{w_i}\, y_i}^{y'} - a \overbrace{\sqrt{w_i}\, x_i}^{x'} - \overbrace{\sqrt{w_i}\, b}^{b'})^2$$

$$x'_i = \sqrt{w_i}\, x_i$$

$$y'_i = \sqrt{w_i}\, y_i$$

$$b' = \sqrt{w_i}\, b$$

$$\min_{w,b} \sum_i w\, \epsilon_i^2 = \min_{w,b} \sum_i (y'_i - a x'_i - b')^2$$