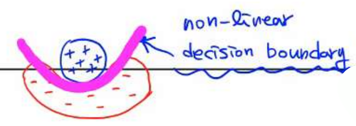
Locally Weighted Logistic Regression

Wednesday 22 May 2024 9:37 AM

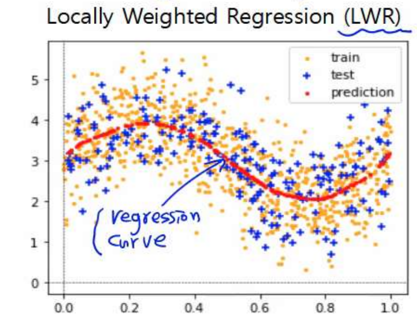


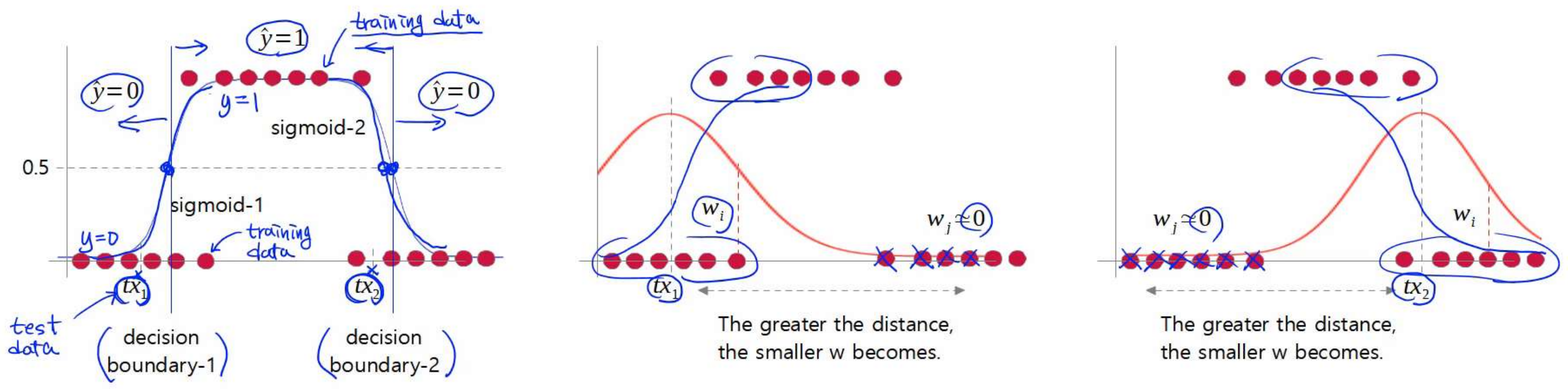
Locally Weighted Logistic Regression (LWLR)

- The concept of LWLR is very similar to Locally Weighted Linear Regression, LWR. LWR can generate a non-linear regression curve, and LWLR can generate a non-linear decision boundary.
- It appears that two sigmoid functions are needed to classify the data in the left picture below. This will give you two decision lines. If tx1, the test data point, is on the left area, it can be predicted using the sigmoid-1, otherwise, it can be predicted using the sigmoid-2.
- Calculate the weight of each data point using a normal distribution in the same way as Locally Weighted Linear Regression, LWR. When tx is on the left, the weights of the data points in the left area are large, so Logistic Regression is performed primarily using those data points.
- This is also a lazy learner.

* reference: [MXML-3-05]

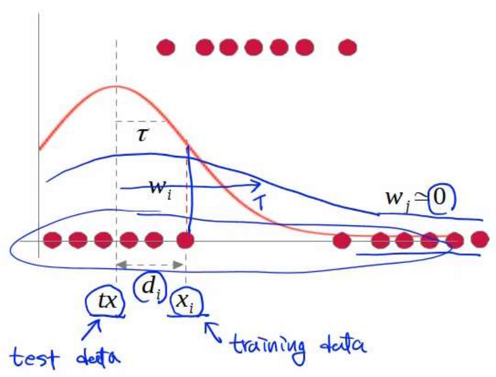
Locally Weighted Regression (LWR)





Weights and Objective function

- Normal distribution is often used as a weight function (or kernel function).
- Calculate the distance d between the test data point tx and all training data points x, and calculate the weight w of each data point using the normal distribution for d.
- The closer the training data point is to tx, the larger the weight and vice versa.
- The standard deviation, tau of the normal distribution can be used to adjust the range of neighbors. The tau is a hyper-parameter.
- Weighted binary cross-entropy is used as the objective function.



• Weights

$$d_i = |tx - x_i|$$

two-dimensional distance

$$d_i = \sqrt{(tx_1 - x_{1,i})^2 + (tx_2 - x_{2,i})^2}$$

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

standard deviation

• Objective function for LWLR

$$\min_{w,b} \sum_i w_i [-y_i \cdot \log \hat{y}_i - (1-y_i) \cdot \log(1-\hat{y}_i)]$$