Other Types of Regression

Till now we have seen, linear regression, where we can write Y, in a linear relationship with X.

In this section we will briefly discuss other types of regression:

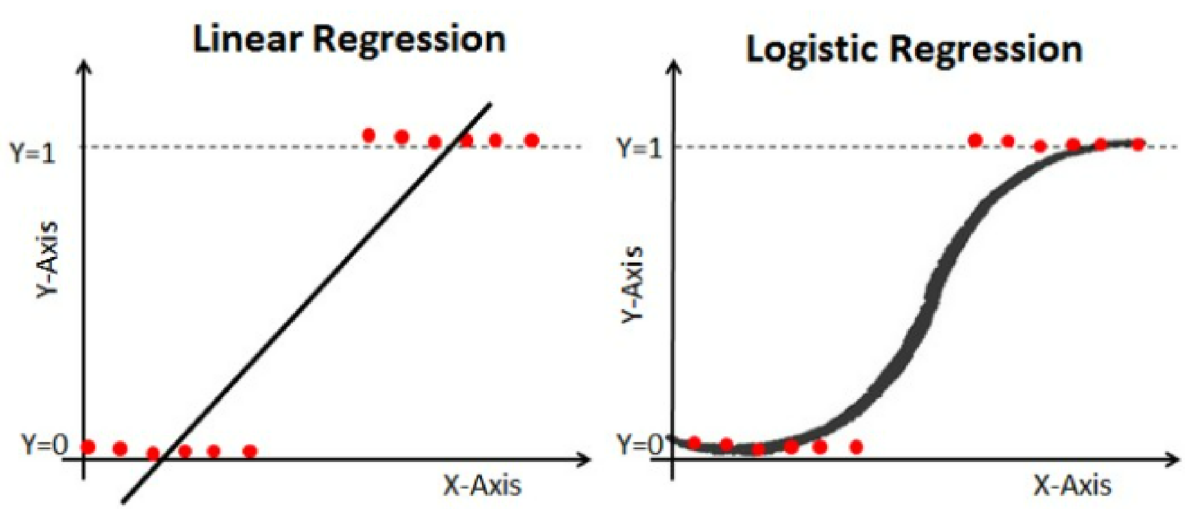
- Regressions induced by using non-linear functional forms, for example, logistic regression.
- Regressions induced by using different types of loss functions instead of the squared loss function, for example, median and quantile regression.

In non-linear regression, we use a non-linear function of X, p(X, β), to predict Y. For instance, the **logistic regression** employs:

$$p(X,\beta) = \exp(\beta' X)/(1+\exp(\beta' X)).$$

- This functional form is particularly attractive when the response variable Y is binary, taking values of 0 and 1, so that the predicted values are naturally constrained between 0 and 1.
- The problem of predicting binary Y is a basic example of a classification problem. We can interpret the predicted values in this case as approximating probability of Y = 1

The graph for the logistic regression looks like as shown below:



As you can see, Logistic regression always lies between 0 and 1. We will learn more about Logistic regression in the next week.

As we did in linear regression, to define an objective function and use it to find the best regression parameter β, here also we can estimate parameters either by solving the sample nonlinear least-squares problem,

$$\min_{b \in \mathbb{R}^p} \mathbb{E}_n (Y_i - p(X_i, b))^2,$$

In the case of binary outcomes, by maximizing the logistic log-likelihood function,

$$\max_{b \in \mathbb{R}^p} \mathbb{E}_n \{ Y_i \ln p(X_i, b) + (1 - Y_i) \ln(1 - p(X_i, b)) \}.$$

We previously used the squared error or squared residuals to set up the best linear predictor.

$\mathbf{E(Y - \beta'X)^2 = \frac{1}{N} \Sigma((Y - \beta'X)^2)}$

We can call this approach the *linear mean regression* because the best predictor of Y under squared loss is the conditional expectation of Y given X.
What if we use the absolute deviation error instead? Then we will obtain the median regression because we will implicitly try to estimate the median value of Y conditional on X.
We define the linear median regression by solving the best linear prediction problem under the absolute deviation loss:

$$\min_{b \in \mathbb{R}^p} \mathrm{E} |Y - b'X|.$$

In the population, we use the population expectation E and in the sample, we use the empirical expectation En instead.

Median Regression

- Just like mean regression, median regression can also be made nonlinear, so we can replace *b*'X by p(X, b).

- Median regressions are great for cases when the outcomes have heavy tails or outliers. They much better behave in this case than the least-squares regression.

Quantile Regression

If we use an asymmetric absolute deviation loss, then we end up with the asymmetric absolute deviation regression or quantile regression.

We know the Linear regression model equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

Where p is the number of regressor variables n is the number of data points.

The best linear regression line is found by minimizing the mean square error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\right)^2$$

Quantile Regression Model Equation for the τth quantile is:

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_p(\tau)x_{ip}$$

Where p is the number of regressor variables n is the number of data points.

The best Quantile regression line is found by minimizing by minimizing median absolute deviation.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - (\beta_0(\tau) + \beta_1 x_{i1}(\tau) + \cdots + \beta_p(\tau)x_{ip})\right)$$

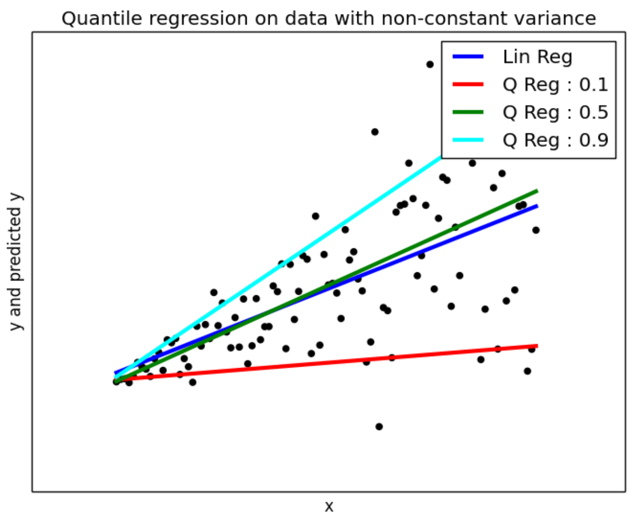
Here the function ϱ is the check function which gives asymmetric weights to the error depending on the quantile and the overall sign of the error. Mathematically, ϱ takes the form

$$\rho_\tau(u) = \tau \max(u, 0) + (1 - \tau)\max(-u, 0)$$

Specifically, here we define the linear τ-quantile regression by solving the best linear prediction problem under the asymmetric absolute deviation loss:

$$\min_{b \in \mathbb{R}^p} \mathrm{E}\left(\tau |Y - b'X| 1(Y > b'X) + (1-\tau)|Y - b'X| 1(Y < b'X)\right)$$

In the population, we use the population expectation and in the sample, we use the empirical expectation En instead.
The weight τ, which appears in the definition of the absolute deviation loss, specifies the τ × 100-th percentile of Y that we are trying to model.



Quantile regression is great for determining the factors that affect the outcome in the tails. Examples:

1. In risk management, we might forecast the extremal conditional percentiles of Y using the information X; this is called conditional value-at-risk analysis.
2. In medicine, we could be interested in how smoking and other controllable factors X affect very low percentiles of infant birth weights Y.
3. In supply management, we could try to predict the inventory level of a product that is able to meet the 90-th percentile of demand Y given the economic conditions described by X.