LinearRegression

February 3, 2019

1 Linear Regression

- Regression analysis helps us to understand how much the dependent variable changes with a change in one or more independent variables
- Forecast or impact of chages. Identify the strength of the effect that the independent variable(s) have on a dependent variable.
- Predict trends and future values.

1.0.1 Simple Linear Regression

Equation of line:
$y = w_1 x_1 + w_2$
where,
slope: w_1
y-intercept: w_2

Error Functions The two most common error functions for linear regression are: - Mean Absolute Error (MAE) 2. Mean Squared Error (MSE)

Mean Absolute Error:

- The vertical distance from the point to the line is the $y - \hat{y}$.

Mean Absolute Error is the sum of all the absolute errors divided by the number of points:
$Error = \frac{1}{m} \sum_{i=1}^{m} |y - \hat{y}|$
Using gradient descent we get the best possible fit line with the smallest possible MAE.

Mean Squared Error: Mean Squared Error is the sum of all the squared errors divided by the number of points:
$Error = \frac{1}{2m} \sum_{i=1}^{m} (y - \hat{y})^2$
By minimizing the average sum of squared errors, MSE is minimized and we get the best possible fit line.

Mean Squared Error **Data:**
$x_1, x_2, ..., x_m$
Labels:
$y_1, y_2, ..., y_m$
Predictions:
$\hat{y}_i = w_1 x_i + w2$
where,
slope: w_1
y_intercept: w_2
Mean Squared Error:
Given the values of w_1 and w_2, we can calculate the predictions and the error based on these values of w_1 and w_2.
$Error(w_1, w_2) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$
In order to minimize this error, we need to take the derivatives wrt the input variables w_1 and w_2 and set them both equal to zero. We calculate the derivites and we get these two formulas:
$0 = \frac{\sum {x_i}^2}{m} w_1 + \frac{\sum x_i}{m} w_2 + \frac{\sum x_i y_i}{m}$
$0 = \frac{\sum x_i}{m} w_1 + w_2 + \frac{\sum y_i}{m}$
Now, we need to solve for w_1 and w_2 for these 2 equations to be zero. We have a system of two equations and two unknowns to be solved.
For a system with greater number of dimensions in inputs, the problem would have n equations with n unknowns which will need a lot of computational power depending on the size of n.
Therefore, gradient decent method is used to obtain a solution that fits our data very well.

1.0.2 2-Dimensional solution

Data:
$x_1, x_2, ..., x_m$
Labels:
$y_1, y_2, ..., y_m$
Predictions:
$\hat{y}_i = w_1 x_i + w_2$
where,
slope: w_1
y_intercept: w_2
Mean Squared Error:
Given the values of w_1 and w_2, we can calculate the predictions and the error based on these values of w_1 and w_2.
$E(w_1, w_2) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$
We need to minimize this error function. Ignoring $\frac{1}{m}$, and replacing the value of $\hat{y}_i = w_1 x_i + w_2$, we get:
$E(w_1, w_2) = \sum_{i=1}^{m} (\hat{y} - y)^2 = \sum_{i=1}^{m} (w_1 x_i + w_2 - y)^2$
In order to minimize this error function, we need to take the derivatives wrt w_1 and w_2 and set them equal to 0.
Using the chain rule, we get:
$\frac{\partial E}{\partial w_1} = \sum_{i=1}^{m} (w_1 x_i + w_2 - y_i) x_i = w_1 \sum_{i=1}^{m} {x_i}^2 + w_2 \sum_{i=1}^{m} x_i - \sum_{i=1}^{m} x_i y_i$
and