1. Explain the linear regression algorithm in detail.

Answer -->
Linear Regression is Supervised Machine Learning Algorithm which find the best linear fit relationship between dependent and independent variables.

The best fit line is created by minimizing sum of residual squares, i.e. sum of squares of difference between predicted and actual values.

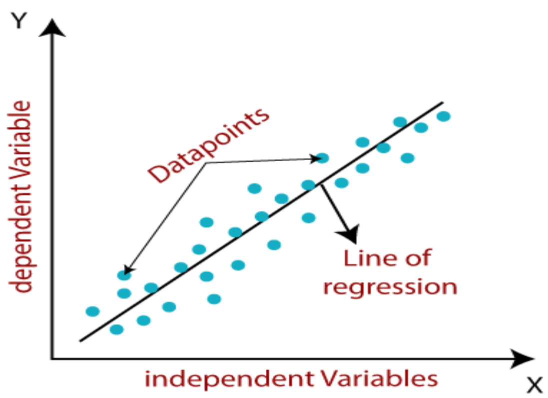
$$y=\beta_0X_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n+\ \Omega$$

β – weights/coefficients/feature_gradient

Ω - bias/intercept/default_value

- **weights-** gradient value that signifies, change in the target value with unit change in one of the feature variable when all other variables are held constant.
- **bias** – default value when all the features are not considered and are insignificant.

Basic Assumptions: before using this algorithm are:

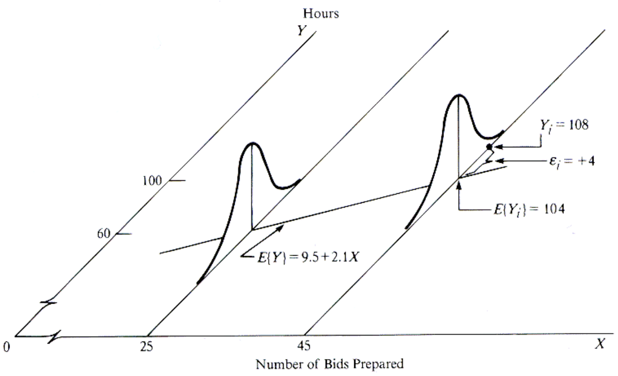


1. almost linear relationship between target and each individual feature variables, checked by two-dimensional pairwise scatter/line plot.

2. feature variables that are recorded without any error, are independent of each other, i.e. no multi-collinearity in our data.

3. error terms:

FIGURE 1.6 Illustration of Simple Linear Regression Model (1.1).



- normal distribution of each error terms.
- mean of error terms is 0, as we have both equal numbers of positive and negative errors.
- error terms are independent of each other, no correlation between the error terms, as our feature variables are independent.
- all error terms must have equal standard deviation.
- Error terms plot should not represent a pattern, random values for error terms.

$$Y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$$

Where Y = Dependent Variable(DV)

X_1, x_2, x_n – Indepdent Variable(IV)

b_0 – intercept

b_1, b_2 – coefficients

n – No. of observations

Hypoothesis Testing: we assume a null hypothesis that the features are insignificant, thus coefficients of these features in the equation are 0.

Thus when we study on model, we infer about the significance of the features by:

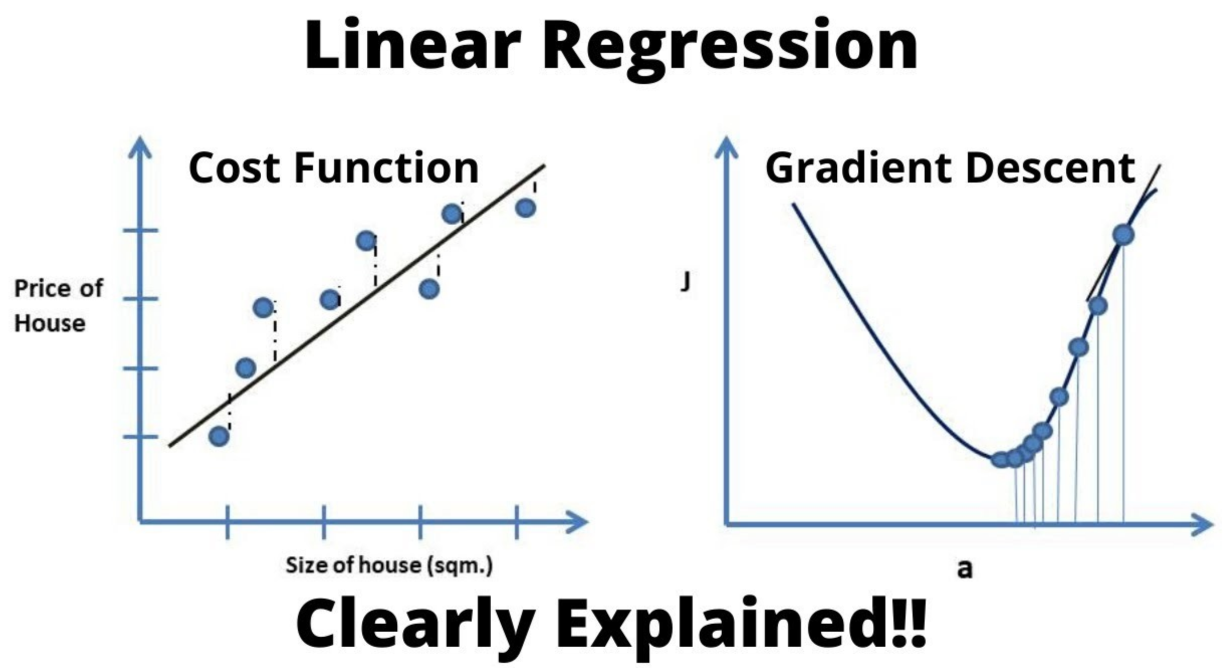
- p-value: if value is less than the significance level, i.e. 0.05 (default), the feature is significant, otherwise it's insignificant.
- Coefficient: if the coefficient value is 0, the variable is insignificant.
- VIF: if the variation inflation factor of the value is very high, we remove those features and check our all values again.
- Adjusted R^2: if the value is very less, we remove some variables which are not logically correct to use in our model.

Cost Function: we determine the values of the coefficients using this method. Cost function is basically mean square error of the function.

$$Cost\ Function(MSE) = \frac{1}{n}\sum_{i=0}^{n}(y_i - y_{i\,pred})^2$$

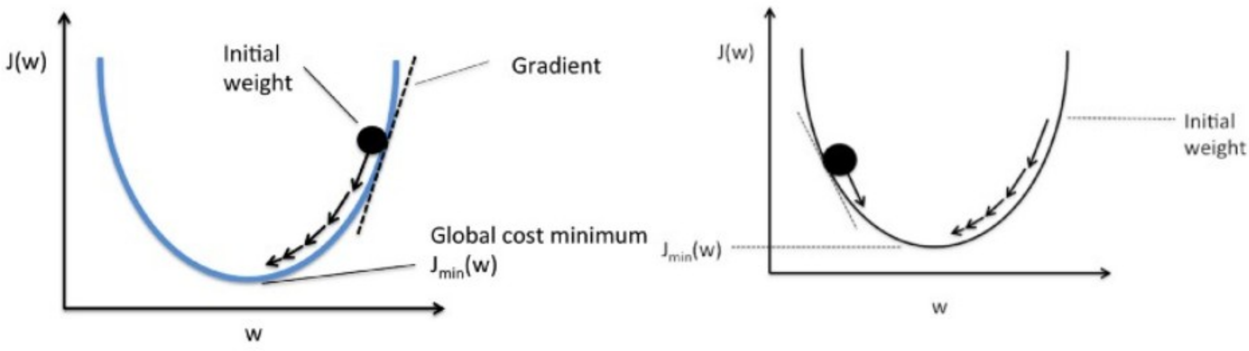
Replace $y_{i\,pred}$ with $mx_i + c$

$$Cost\ Function(MSE) = \frac{1}{n}\sum_{i=0}^{n}(y_i - (mx_i + c))^2$$



gradient descent algorithm: In linear regression model, we normally use the iterative algorithm to find th eour best possible values of coefficients and intercept.
Usually the best possible value is obtained at the global minima of the cost function curve which is also called convex curve.
We choose random values for the coefficients and intercepts, e.g. 0 & 0, and a very small learning rate (a step

size value for iterations) to move to the next values.



The equation for the next values of coefficients and intercept is the difference in the initial value and the learning rate times the gradient with respect to the specific weight/bias at the initial value.

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left((h_\theta(x_i) - y_i)x_i\right)$$

If our initial value is greater than the minima, the gradient will be positive and the new value will be less than initial value. Otherwise if the initial value is less than the minima, gradient will be negative and thus the sign will reversed in the equation and the new value will be more than the initial value.

Using this iterative approach, we reach towards the optimal values of the coefficients and intercepts where cost function is minimum globally.

Model Training:

1. Split our dataset into training, validation and test dataset.
2. Fit our model using training dataset, our models learns specific values like mean, standard deviation and distribution from the dataset, and calculates coefficients and intercepts for the best fit line using gradient descent method.
3. Use validation dataset to validate our model and predict values on the validation dataset and calculate r2 score.
4. Use the model to predict values on the test dataset, calculate r2 score for the predicted values and actual values in the test dataset.
5. Compare r2 score for validation and test dataset and check for under-fitting or over-fitting.
6. If some major discrepancy found, do tuning (coarse-fine) on the dataset using recursive feature elimination and manual testing.