Insights of Machine Learning Algorithms.

PART-1

We'll Discuss The Insights of These Algorithms.
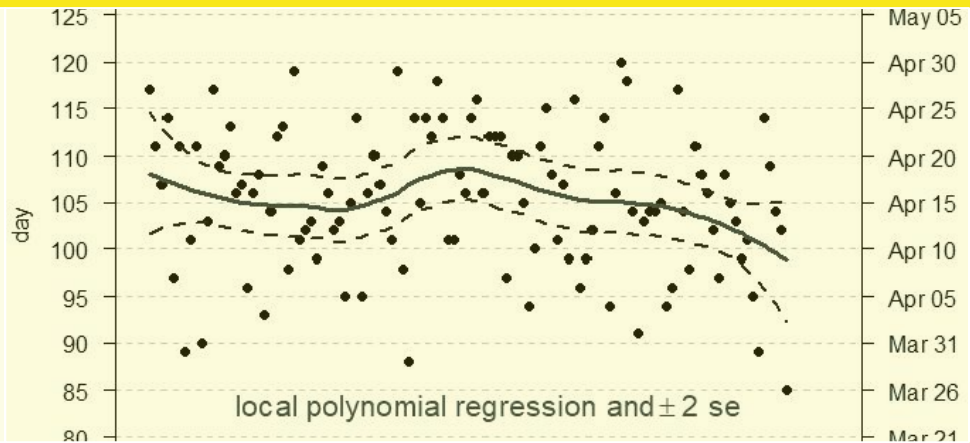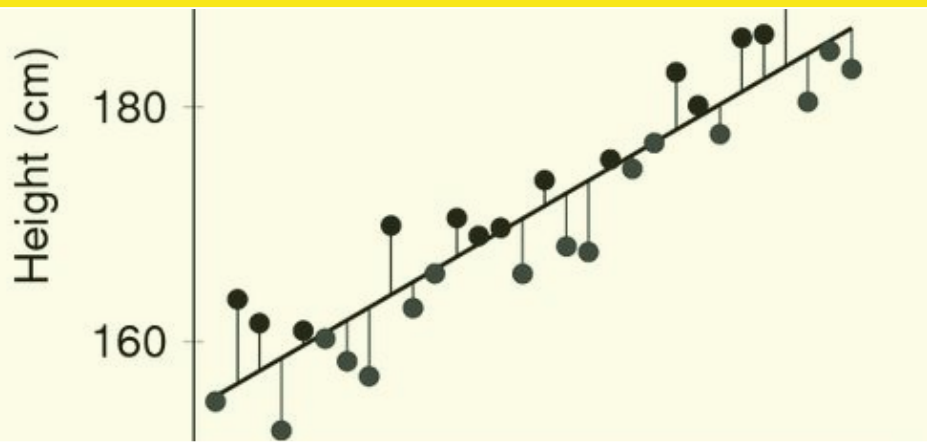
- ❏ Linear Regression
- ❏ Logistic Regression

Let's Begin Our Journey!!!

NOTE!!

❖ Here I'm not going to discuss mathematics Behind the algorithm. Instead of that I'm going to share the key insights behind that. Which tells you how & why we can apply any algorithm for the predictive analysis.

Linear Regression!!

local polynomial regression and ±2 se

Linear Regression.

1. It comes under supervised learning algorithm that means finding the O/p based on the inputs relations.
2. As it's name suggested Linear Regression the relation between I/p & O/p must be linear.
3. Therefore in LR problem we deal with linear data. That can be separated by straight line.
4. The output is continuous in nature, that means every time we will get different O/p based on its I/p.
5. We can apply this algorithm when there is correlation between I/ps & O/p.
6. Don't confuse with multicollinearity. Which means I/p variables are highly correlated which each other which means they aren't going to give new information we can use one of them.

❖ **Linear Regression is of three types:**

➤ Simple Linear Regression.

   ■ It is the regression technique in which one I/p & one O/p is there.

   ● $Y = B_0 + B_1X$

➤ Multilinear Linear Regression.

   ■ It is the regression technique in which more than one I/p & one O/p is there.

   ● $Y = B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n$

➤ Polynomial Linear Regression.

   ■ It is the regression technique in which we deal with one I/p & one O/p variable where we take exponential nth degree of polynomial of X.

   ● $Y = B_0 + B_1X_1 + B_2X_1{}^2 + \ldots + B_nX_1{}^n$

Tips for fun!!!

1. Polynomial Regression is the regression which deals with non-linear data.
2. While working with LR keep in mind to avoid outliers.
3. All the I/p variables should be numeric in nature.
4. Good practise to standardize the I/p fields so all the values will be in same range which will reduce the error,
5. MSE And RMSE should be low they will give the intuition of error present in the model. If it is too low it'll underfit the data or if it is very high it'll overfit the data.
6. R2 should be high it will tells you how best the line fits the data. Which assumes that every single variable explain the variation in the dependent variable. Therefore the value of R2 will increase with increase in I/p variable.
7. Preferable is Adjusted R2 because it tells you the percentage of variation explained by the I/p variables that actually affect the dependent variable . It'll not increase with increase in I/p variable until it really affects the O/p variable.

Logistic Regression!!

Logistic Regression!!

1. It also comes under Parametric Supervised learning Algorithms.
2. Though logistic regression or logit regression is a linear method but the predictions are transformed using logistic function.

$$P = 1/1+e^{-z}$$

Or

$$p/1-p = e^{B0+B1X}$$

3. Therefore, softmax function is used to classify the binary distribution.
4. This function can't give the clear insight that if there is change in X or in B1 what will be the change in probability.
5. To overcome this problem we introduced the concept of odd logs.

1. So we're going to take log both sides and made the exponential equation linear.

$$\ln(p/1-p) = B0 + B1X$$

❖ Thus, the log odds & odds are widely used to find the linear relationship in the field of Finance, Marketing, Telecom Industry, etc.

❖ Logit Regression are of three types.

➢ Binary Logit Regression:

■ It contains of binary classification(0,1) at target variable side.

➔ Multinomial Logit Regression:

◆ Three or more categories without ordering will be applicable at target variable side.

➔ Ordinal Logit Regression:

◆ Three or more categories with ordering will be applicable at target variable side.

Tips for Fun!!!

1. It gives convenient probability scores for observations.
2. Multicollinearity is not an essential issue & it can be countered by L2 regularization.
3. Although, it's less prone to overfit but with high dimensional dataset it can be overfit to overcome this problem we should use L1 & L2 Regularization techniques.
4. All the fields must be numeric in nature no categorical feature is allowed.
5. Relies on transformation for non-linear features.
6. Normalization will increase the accuracy of the result.
7. It is prone to outliers so avoid if there is any.
8. Data cleaning and filling missing values is the part of data processing which is needed in this.