**Machine Learning Notes**

Regression

* Regression models (both linear and non-linear) are used to predict a real-life value, like salary for example.
* If your independent variable is time, then you are forecasting future values, otherwise your model is predicting present but unknown values.
* Regression techniques vary from Linear Regression to SVR and Random Forest Regression.
* The linear regression included in this study are –

1. Simple Linear Regression
2. Multiple Linear Regression
3. Polynomial Regression
4. Support Vector for Regression (SVR)
5. Decision Tree Regression
6. Random Forest Regression

**Simple Linear Regression**

* A simple linear regression is basically the formula **y = b0 + b1\*x1**, which is basically the formula of trend line or sloped line on x/y axis.
* In the formula, y = dependent variable (something you are trying to explain, for e.g., how does a person’s salary change over the years). Something you are trying to understand, how it depends on something else.
* X is the independent variable. In simple linear regression we only have one independent variable so we can call it just x. This is the variable that we are assuming that is causing the dependent variable to change.
* b1 is the coefficient of the dependent variable, and it kind of explains how the unit change in x1 affects the unit change in y. It is kind of a connector between y and x1.
* We cannot always say that change in x will always equals to change in y, there might be a coefficient or proportion, to which proportion this change is taking place.
* b0 is a constant term. It means where the point where line crosses the vertical axis.

Chart, scatter chart

Description automatically generated

* In the above example, if we apply the formula of simple regression, we can conclude that **if the coefficient b1 is less then slope will be less, and the salary increase will be less per every year of experience. And if the slope of line is greater, the salary increase will be more per every year of experience.**
* So, the core working of Simple Linear Regression is that we are using observations that we have to find the best fitting line.

**Picking The Best Fit Line**

* To pick the best fitting line, we first take the sum of all the squared differences between the actual value and the predicted value.
* SUM (*y* - )2
* Once you have the sum of the square differences for all the values, we choose the one that has the minimum value.
* So, basically what a simple linear regression does is draw all possible trend lines, calculates the sum of squared differences of the lines, records the calculated sum for all the trend line, and then It finds the minimum sum of squared.
* The above-mentioned method is called **Ordinary Least Squared Method**.

Note – In Simple Linear Regression we will only have one independent variable, one feature, and only one continuous variable to predict.

Regression is when you predict a continuous variable like salary.

Classification is where you classify the variable in a certain class or a category like type of users and such.