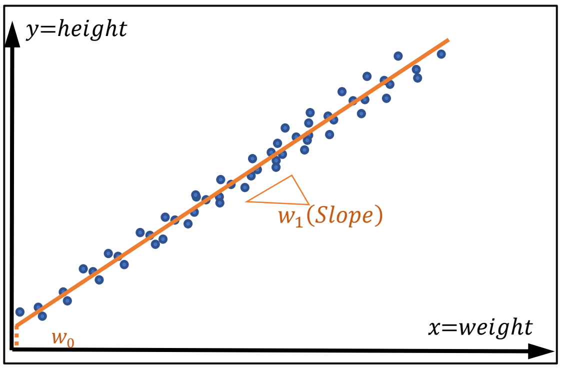
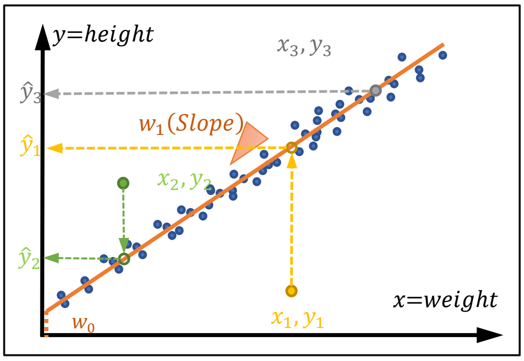
- **Linear Regression**
  - Linear regression models the relationship between two variables by fitting a linear equation to observed data.
  - One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. E.g., We want to relate the weights(kg) of individuals to their heights(m) using a linear regression model.
  - In Linear Regression, we are trying to find a line that fits our data as much as possible



  - Equation Of a line is $\boxed{\begin{array}{c} y = m * x + c \\ Height \ = w_1 * Weight + w_0 \end{array}}$
  - Similar Concept can be extended to 3D where for pair of features $f1$ & $f2$. Here, we have to predict with a plane that best fits the data.

    Equation of a plane $\boxed{\begin{array}{rl} y_i & = w_1 * x_{i1} + w_2 * x_{i2} + w_0 \\ Height & = w_1 * weight + w_1 * age + w_0 \end{array}}$
  - After Linear Regression, we found following line which fits the data well



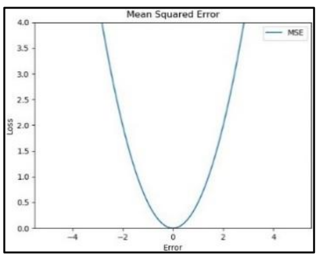
  - There are 3 query points x_1, x_2, x_3 But two points x_1, x_2 are wrongly predicted as they do not fit the line.

$$\boxed{\begin{array}{l} error_{x1} = \hat{y}_1 - y_1 > 0 \\ error_{x2} = \hat{y}_2 - y_2 < 0 \\ error_{x3} = \hat{y}_3 - y_3 = 0 \end{array}}$$

  - The ultimate task in linear regression will be, to minimize the sum errors. Now since there are both positive & negative errors, we have to minimize square of the errors.
  - So, we want to find optimal w & optimal w_0. w is a slope which will be a vector & w_0 will be a scalar.

$$\boxed{\begin{array}{c} (W^*, W_0^*) = argmin_{w,wo} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ \hat{y}_i = W^T x_i + w_0 \\ (W^*, W_0^*) = argmin_{w,wo} \sum_{i=1}^{n} \{y_i - (W^T x_i + w_0)\}^2 \end{array}}$$

- **Loss Minimization**

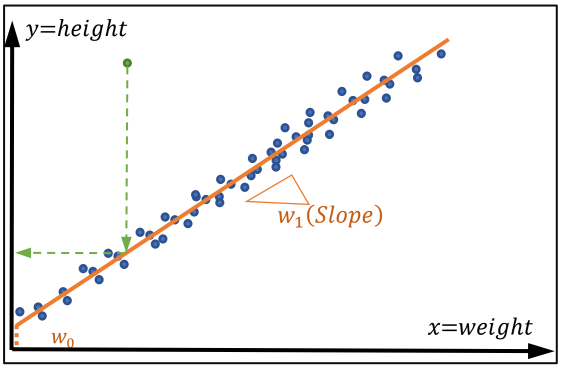


  - Whether the error is positive or negative, we want to penalise it with higher positive value. thus, we are taking square of the error.
  - The problem with $abs()$ function is that it is not differentiable at zero and hence we do not use it.
  - L2 Regularization

$$\boxed{(W^*, W_0^*) = argmin_{w,w_0} \sum_{i=1}^{n} \{y_i - (W^T x_i + w_0)\}^2 + \lambda \left|\left|W\right|\right|_2}$$

  - Similarly, we can use L1 Regularization too

- **Need of regularizer in linear regression**
    - For logistic regression we introduced regularization to constrain weight elements reaching infinity.
    - In real world we round up the values to an easily understandable decimal; this results in small errors even though we know the underlying relationships.
    - Without regularization we will have noise propagated to the weight values; to reduce the effect of noise in features on the weight vector we introduce regularization
- **Outliers**



    - Outliers increase the squared loss heavily
    - So first we compute w* & w0* and find out the points which are far away from the plane and we remove this point from dataset
    - Then we repeat the same procedure on remaining data
    - This technique of iterated removal of outliers from model training is called RANSAC