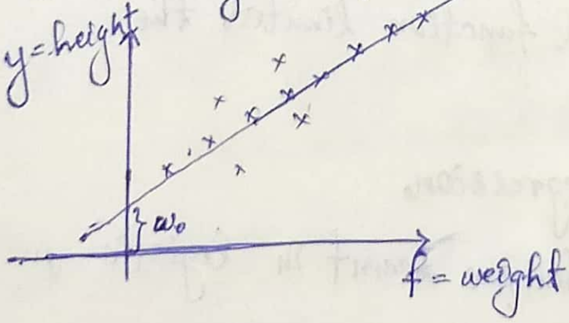
Linear Regression

↳ does actual regression.

Ex → predict height $\in \mathbb{R}$ given weight, gender, ethnicity, hair color, etc.

consider only one feature → weight for simplicity.



Linear regression → find a line that best fits the given data.

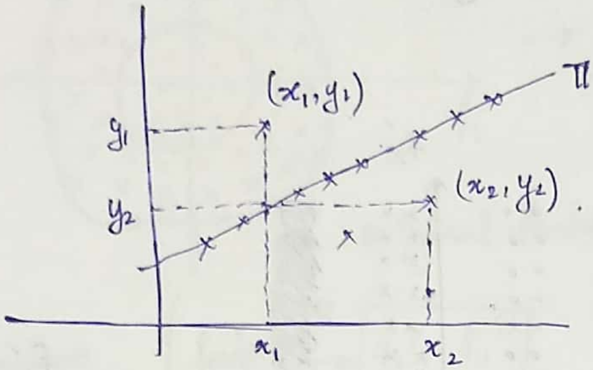
$$\text{height} = w_1 * \text{weight} + w_0$$

$$\underline{y = mx + c}$$

In general, we have to predict a hyperplane that best fits the given data.

Now, what does best-fit mean?

→ let's say we have got the model — $\pi_\circ(w_1, w_0)$.



let the predⁿ by model be denoted by $\hat{y}$.

So, for the model to best fit, the error must be minimum i.e. diff b/w predⁿ & actual values

This must be min. for best fit. ← $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ — squared errors

Mathematical formulation

$$\boxed{(w^*, w_0^*) = \underset{w, w_0}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \text{where } \hat{y}_i = w^T x_i + w_0}$$

with regularization, we have

$$(w^*, w_0^*) = \underset{w_0, w_0}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \|w\|_2^2 \qquad \hat{y}_i = w^T x_i + w_0$$

$$(w^*, w_0^*) = \underset{w, w_0}{\operatorname{argmin}} \sum_{i=1}^{n} [y_i - (w^T x_i + w_0)] + \lambda \|w\|_2^2$$

Real world cases

Almost same as that of logistic regression.

Outliers → In logistic regression, sigmoid function limited the impact of outliers.

But nothing such here in linear regression.

So, we making use of filtering technique learnt in logistic regression, filter out outliers.