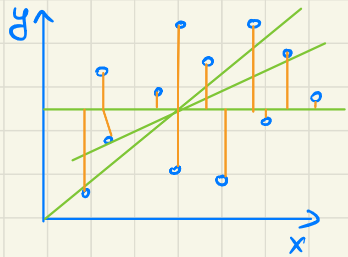
LINEAR REGRESSION

Main Ideas:

1. Use least squares to fit a line to the data : $y = mx + b$
2. Calculate R^2
3. Calculate the p-value for R^2



data point - line

- find sum of squared residuals for each line
- find the rotation that has the least sum of squares ∴ fits data to the best

Since slope ≠ 0, there exists a correlation b/w the variables

Potential problem: If we only have 2 points, we will always find a line that passes through both points ∴ $R^2 = 1$

Pros:

- works well with linear data
- computationally efficient ∵ we only need to plug in values
- easy to understand and interpret and helpful for noticing relationships b/w variables

Multiple Linear Regression

- more than one independent variable

eg: $\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$

- use case: real estate pricing ∵ property rates are based on factors like location, size, # bedrooms, etc.

$\hat{y} - y$: cost function

In linear regression, MSE is the cost function :

$$J = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

How to optimize the loss function to iteratively reduce the MSE?

↳ Gradient descent

Evaluation metrics

↓ RMSE

↓ R^2

↓ Mean Absolute Error (MAE)