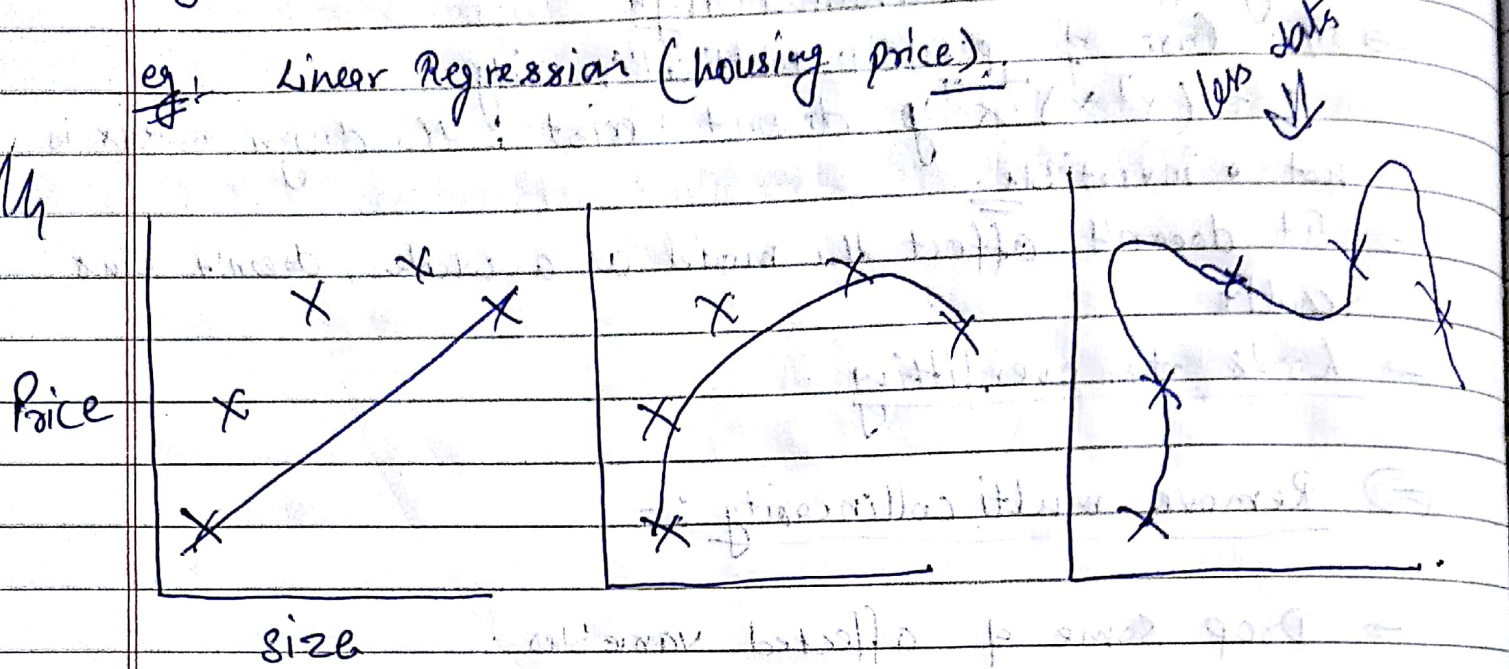
Bias - Variance Trade-off :-

Regularization - Solving the problem of overfitting.

eg: Linear Regression (housing price).



→ $\theta_0 + \theta_1 x$ → $\theta_0 + \theta_1 x + \theta_2 x^2$ → $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

"Underfit" / "high bias" "Just right" "Overfit" / "high variance"

"Large λ" "Intermidate λ" "small λ"

→ Overfitting :- If we have too many features, the learned hypothesis may fit training set very well

$$\left(J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \approx 0 \right)$$

but fails to generalize to new test data.

→ Underfitting : In which the data clearly shows structure not captured by the model.

- Occurs when the form of our hypothesis function is maps poorly to the trend of the data. It is caused :

too simple to hypothesis function or uses too few features

eg:- We have 10 students in classroom. we intend to train a model based on their past score to predict the future score. There are 5 male & 5 female in class. The avg score of male is 80 & female is 60. The overall class avg is 70.

Now several ways to make prediction

(1) Predict the score as 70 for the entire class.

(2) Predict score of male = 80, female = 60. This is simplestic model which might give a better estimate than the first one.

(3) Now, try to overkill the problem. we can use roll no. of student to make the prediction & say that every student will exactly score same marks as last time. Now, this is unlikely to be true & we have reached such granular level that we can go seriously wrong.

(1) underfit (2) optimum fit (3) overfit

⇒ The error emerging from any model can be broken down into 3 components mathematically

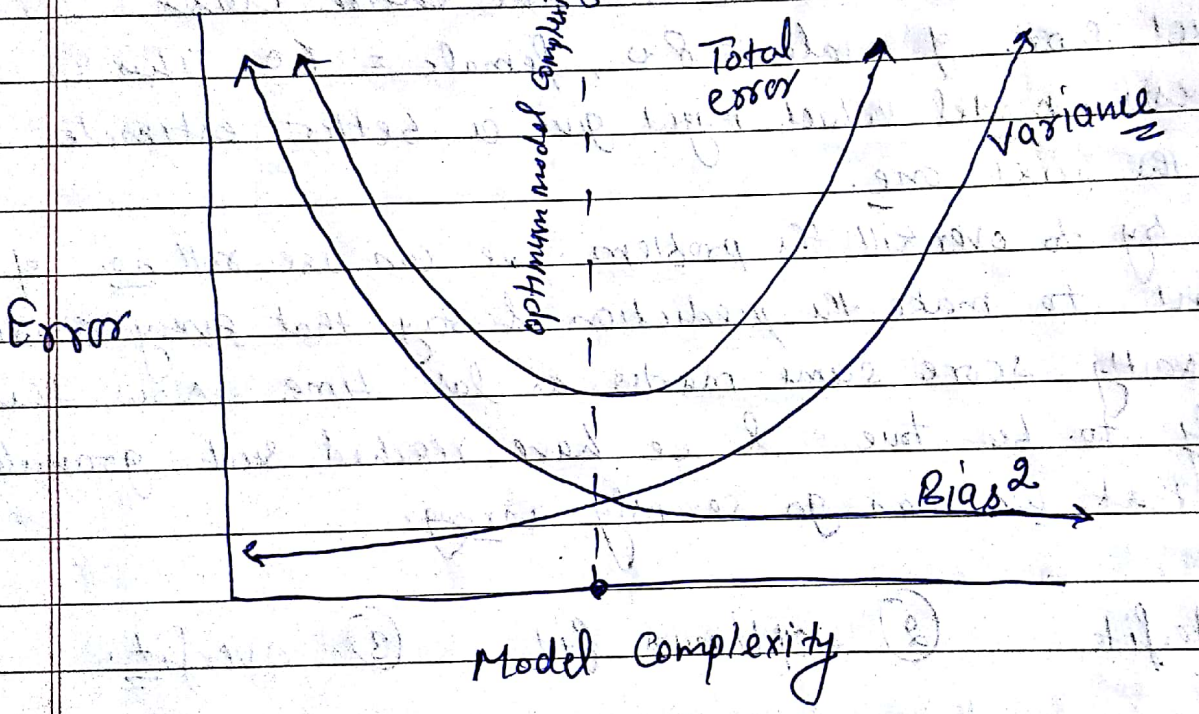
$$Error(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\hat{f}(x) - E[\hat{f}(x)]\right]^2 + \sigma_e^2$$

$$Error(x) = Bias^2 + Variance + Irreducible\ error$$

↑ noise

→ Bias is useful to quantify how much on an avg are predicted values different from actual values.
- A high bias error means we have a under-performing model which keeps on missing important trends.

→ Variance is useful to quantify how are the prediction made on same observation different from each other.
- A high variance model will over-fit on our training data & perform badly on test data.



eg:-
- A person with "High Bias" is someone who starts to answer before even we can finish asking.
- A person with "High Variance" is someone who can think of all sorts of crazy answers
- combining these gives different personalities