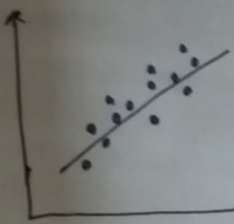
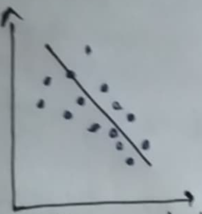
November-20:

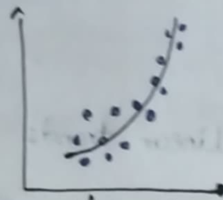
Data → Linear data
Data → Non-linear data.



Linear data



Linear data



Non-Linear data.

* in case of linear data we can perform linear regression, but in case of non-linear data straight line will not be suitable as the distance b/w dots will be more hence, we need to produce curved line as best-fit line.

* Linear data – Linear Regression
  Non-Linear data – Polynomial Regression.

Polynomial Regression: It is an extension of linear regression, where the relationship b/w the x & y is modelled as an nth degree polynomial.

$\hat{y} = \alpha x + \beta$

$\hat{y} = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \dots + \alpha_n x^n + \beta$.

This allows the model to capture non-linear relationship b/w. x & y.

why polynomial regression?
* In Real-world data often doesn't follow a straight line,
* Polynomial Regression helps fit curved pattern such as parabolas, waves or exponential like structure/shapes.

Steps to perform polynomial regression:

1. Choose degree of polynomial (n)
2. Transform features : create $x, x^2, x^3, \dots x^n$.

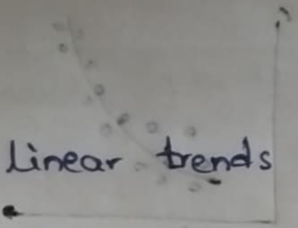
Ex: x_1 x_2 dop: 3

$(x_1, x_1^2, x_1^3, x_2, x_2^2, x_2^3, x_1 x_2$

$x_1^0, x_1^1, x_1^2, x_2^0, x_2^1, x_2^2, x_1 x_2$

3. Fit linear Regression on the transformed feature

4. Evaluate model (R^2 score, MSE, RMSE, ...)



Advantages:

* Captures non-linear trends
* Flexible - can fit many shapes.

Disadvantages:

* High-degree polynomial may lead to overfitting.
* Extrapolation beyond training data is unreliable.
* More complex & harder to interpret.

Usually ML models falls under 2 problems:

1. Overfitting - train accuracy very high, test accuracy very low
2. Underfitting - train & test accuracy are very low.