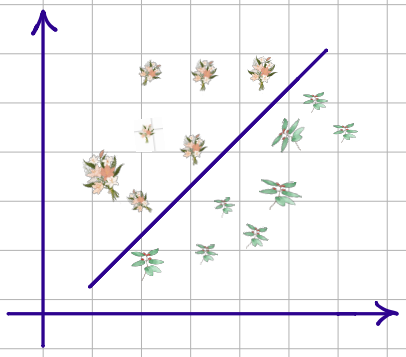
Linear Regression

Type: Supervised

Task: Regression (only)

Steps:

1. Data Preparation → Clean and preprocess
2. Model Definition → Assumes a linear relationship
3. Fit the model
4. Evaluate the model
5. Prediction → on unseen data



Mathematical Foundation:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

- y_i: dependent variable
- $x_1, x_2, \dots, x_n$: independent variable
- β_0: intercept
- $\beta_1, \beta_2, \dots, \beta_n$: coefficients (slope)
- ε: error term

Simple Linear Regression (one-term)

$$y = a + bX$$

- y: dependent variable
- a: intercept
- b: slope
- X: independent variable

Assumptions made:

- Linearity → relationship b/w independent and dependent variables
- Independence → Observations are independent of each other
- Normality of Errors → Errors are normally distributed
- No perfect Multicollinearity → Independent variables are not highly correlated with each other.

Pros

- simple and easy to interpret
- fast to train and predict
- provides quantitative measure of relationships
- useful for feature selection
- diagnostic tool for model assessment

Cons

- assumes linear relationship
- sensitive to outliers
- limited to modeling linear relationship
- assumes independence
- multicollinearity can distort results

How do you handle multicollinearity?

- remove or combine correlated features
- use dimensionality reduction
- apply regularization

What metrics would you use to evaluate?

- MSE
- RMSE
- Mean Absolute Error