Tech Bytes

ML Refresher – Linear Regression
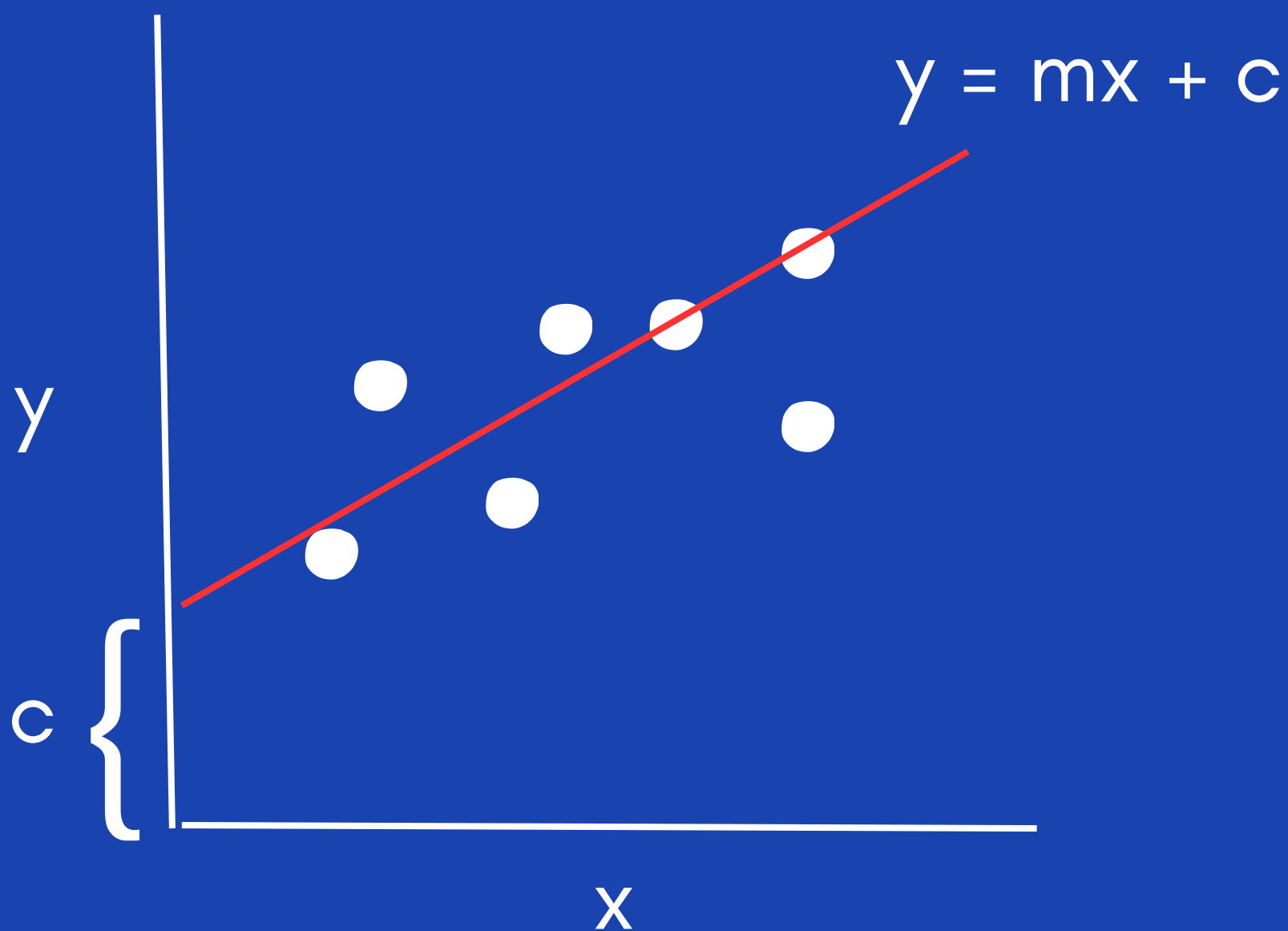
by Pranali Bose

What is it?

Simple Regression model used mostly for interpretability and analysis of relationships between features

Intuition Behind

Finding the best fit line that minimizes the loss function

$$y = mx + c$$

y

c {

x

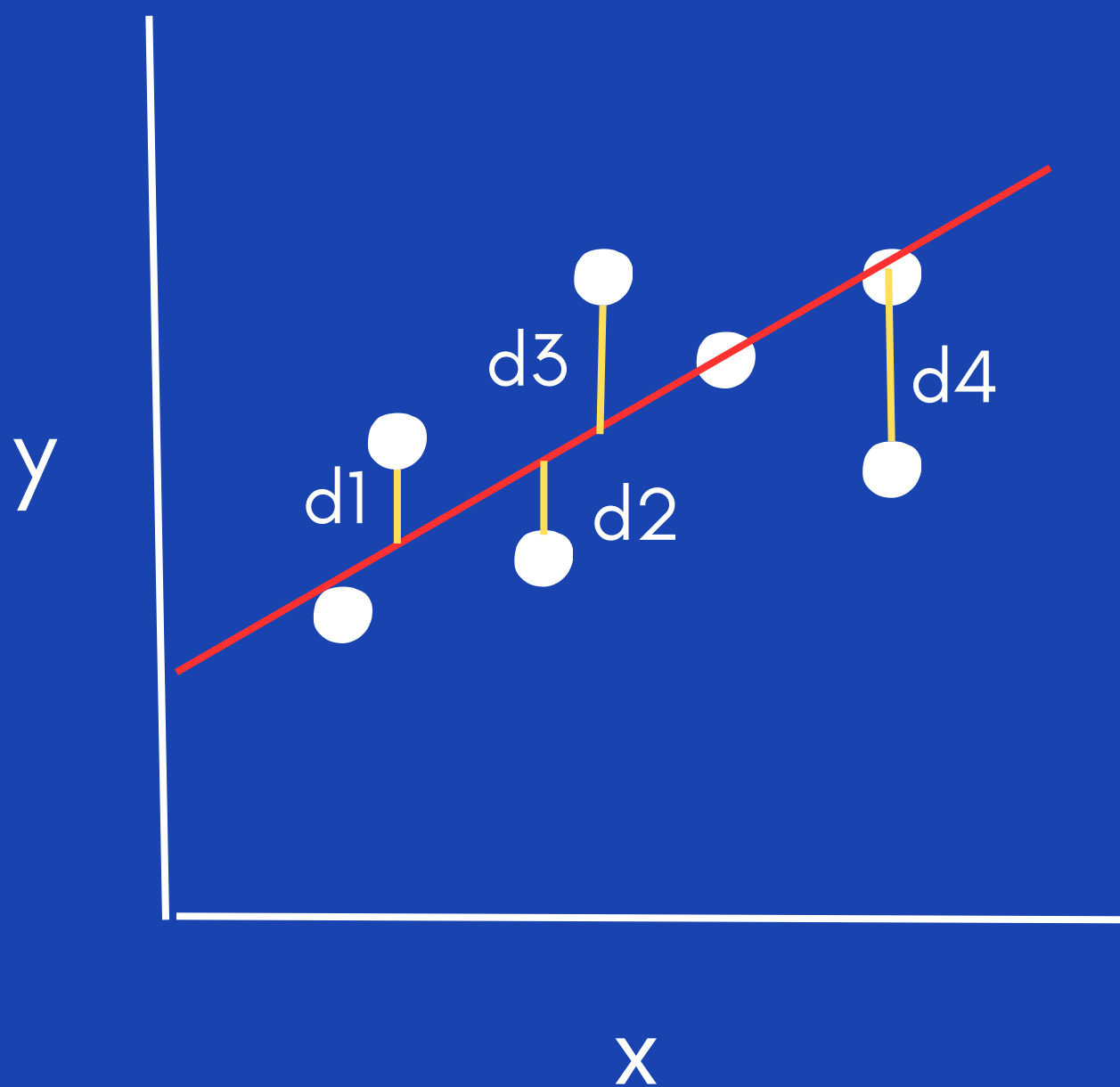x = Independent Variable
y = Dependent Variable
c = Intercept
m = Slope

Loss Function?

Measure of how close the predictions are from the actual values

dn = Difference between actual and predicted values
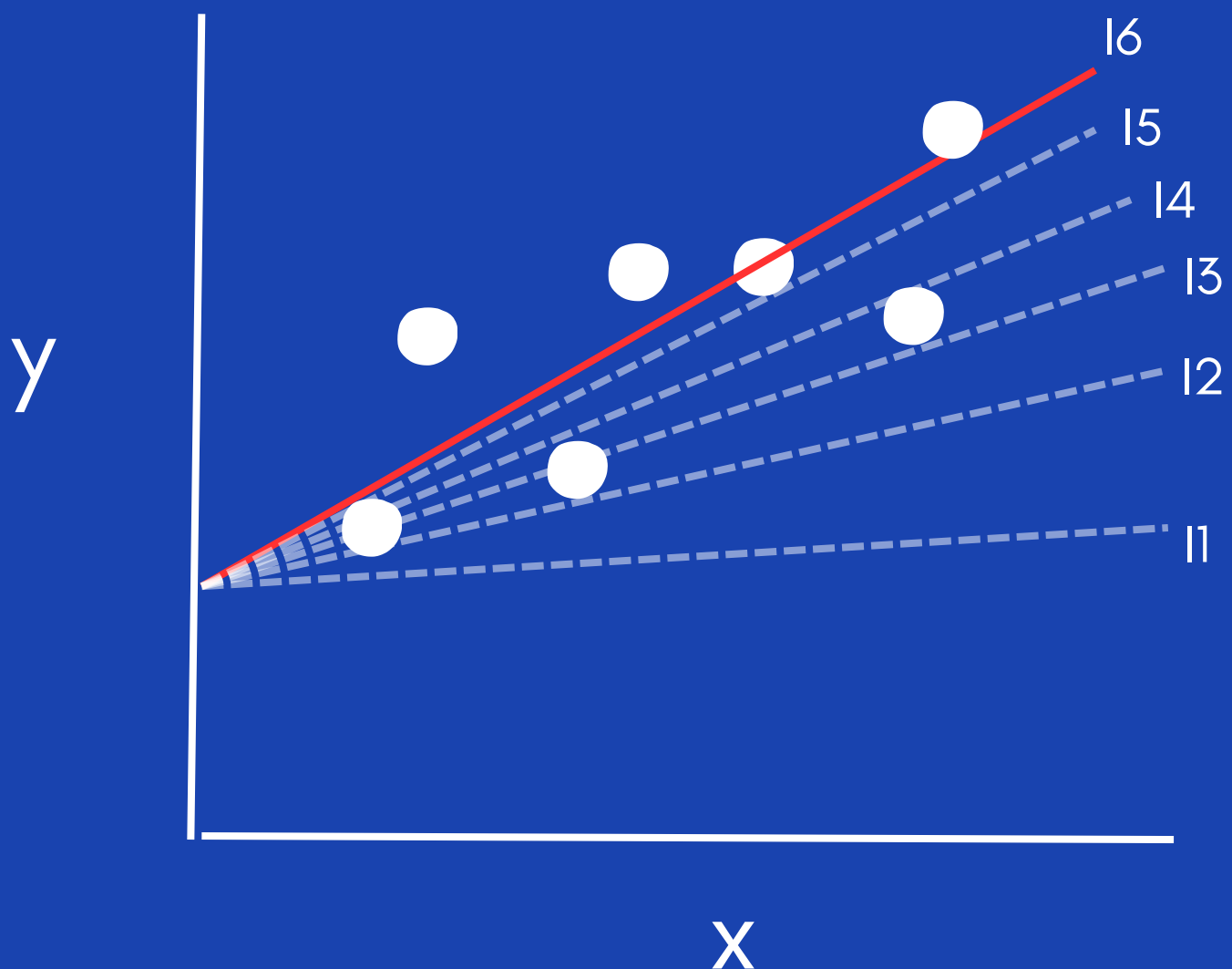
Common Loss Functions

- **L1** or **Mean Absolute Error**
  - $1/N \; \Sigma \; |\text{actual} - \text{predicted}|$
  - Used when dataset has outliers

- **L2** or **Mean Squared Error**
  - $1/N \; \Sigma \; (\text{actual} - \text{predicted})^2$

Training Process

Goal – To find the optimal parameters (m and c)

Process – Iteratively tweak the m and c values and calculate the loss until there is no further improvement (Gradient Descent)

In -> Number of iterations

Model Interpretation

Use Case: House Price Prediction

y = Price (in dollars)
x = Size (in sq ft)

Consider Best Fit Line Equation to be:
$y = 200 * x + 10000$

Intercept (10000): The baseline price of a house when size is zero

Size Coefficient (200): For every additional square foot of size, the price of the house increases by 200

Assumptions

- **Linearity**: The relationship between predictors and the response variable is linear
- **Independence**: Observations are independent of each other
- **Homoscedasticity**: Constant variance of the errors across all levels of the independent variables
- **Normality**: The residuals (errors) of the model are normally distributed
- **No multicollinearity**: Independent variables are not too highly correlated with each other
- **No autocorrelation**: Residuals are not correlated with each other in time series data

Key Hyperparameters

- **Learning Rate**: Controls how much to update the model weights during training

- **Fit Intercept**: A boolean that specifies whether to include an intercept term in the model

- **Max Iterations**: Specifies the maximum number of iterations for convergence in optimization algorithms

Tech Bytes

Find this useful?

Let me know in the comments which topic would you like to see next

Follow for more...