

Today's Agenda:

- Quick Recap
- ML Mind Map
- Preprocessing for ML
- Used Cars Price Prediction
- Rowad Dataset Cleaning and Preprocessing for ML

machine learning

$I_M = 100 * I_{CM} \Rightarrow I_M, 100 CM, \text{rule?}$

Regression

(target)

Continuous

Classification

(tag et)

Categorical

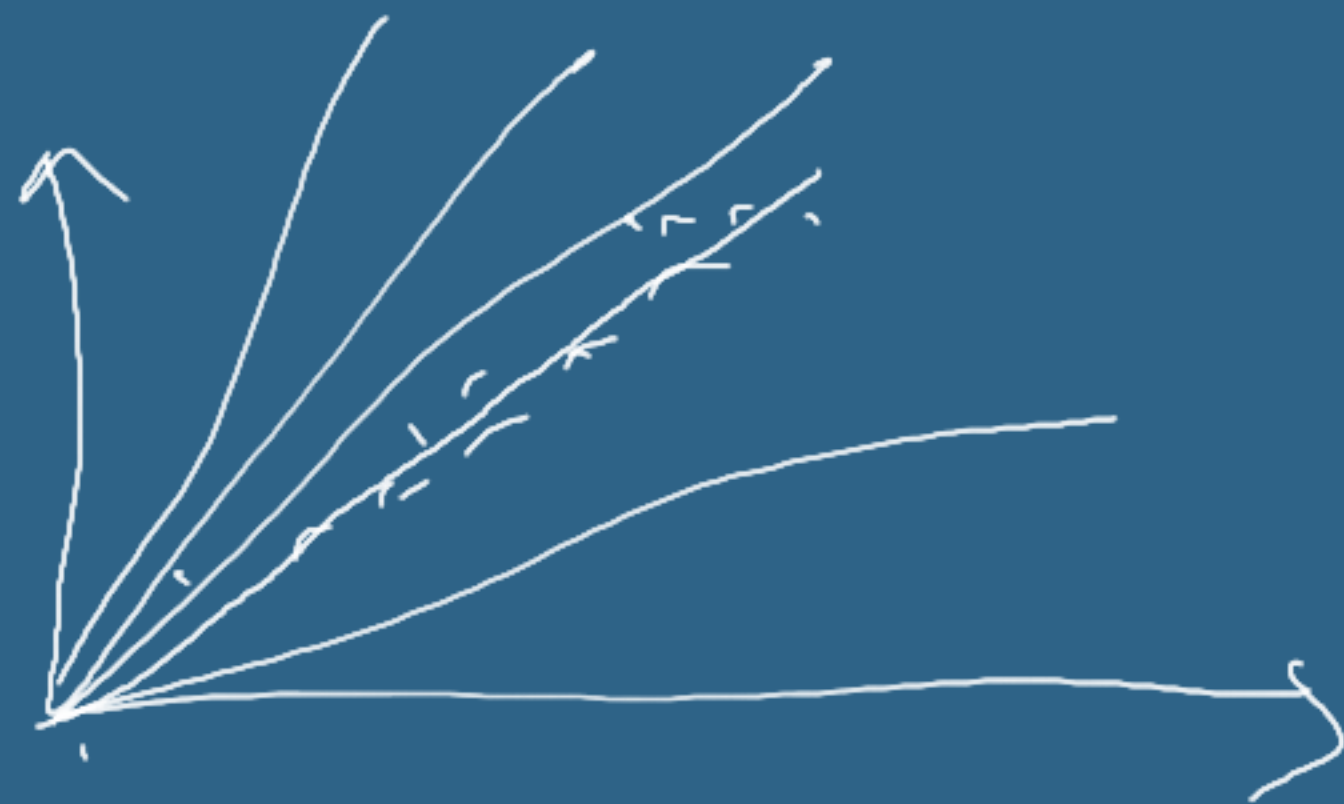
Regression

Simple linear regression

$$y = \theta_0 + \theta_1 x$$

multiple linear regression

$$y = \theta_0 + \theta_1 x + \theta_2 x$$



$$(y - \hat{y})^2$$

Solve for θ (Ordinary least squares)

Analytical Solution
(Normal Equation)

$$\hat{\theta} = [X^T X]^{-1} X^T y$$

Optimization
(Gradient Descent)

$$y = \theta_0 + \theta_1 x$$

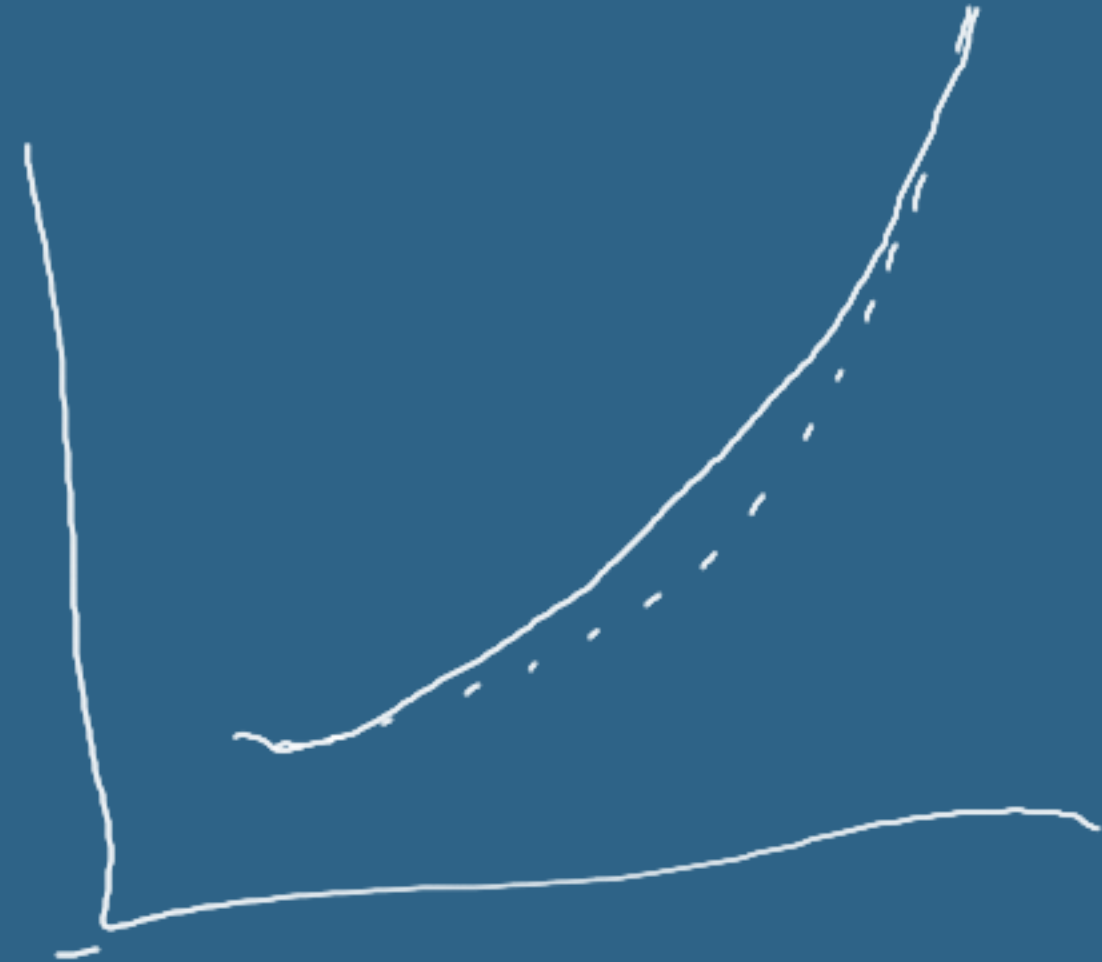
$$J(\theta) = \frac{1}{2n} \sum (y - \hat{y})^2$$

α : learning rate
 $\frac{\partial}{\partial \theta} J(\theta)$: directions

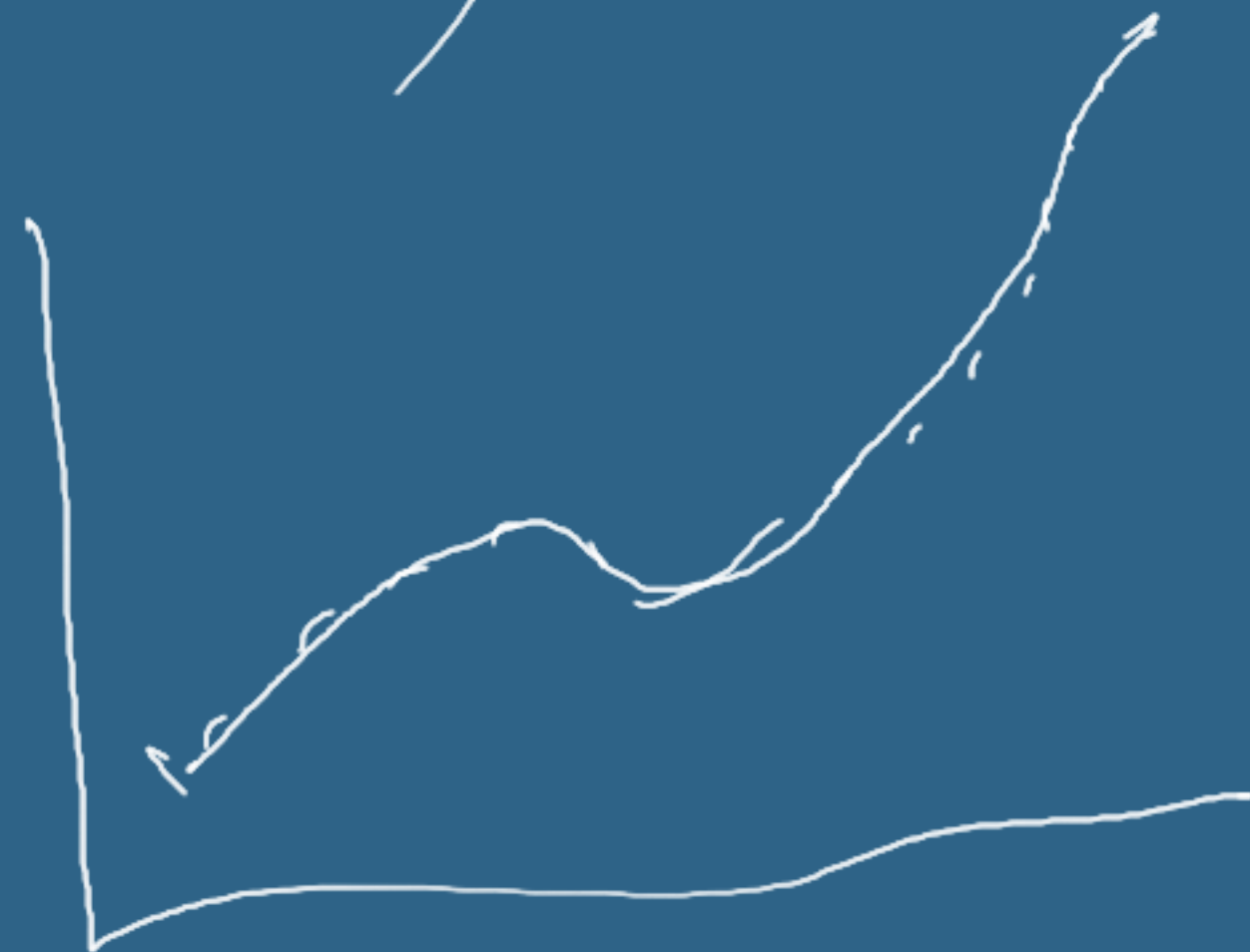
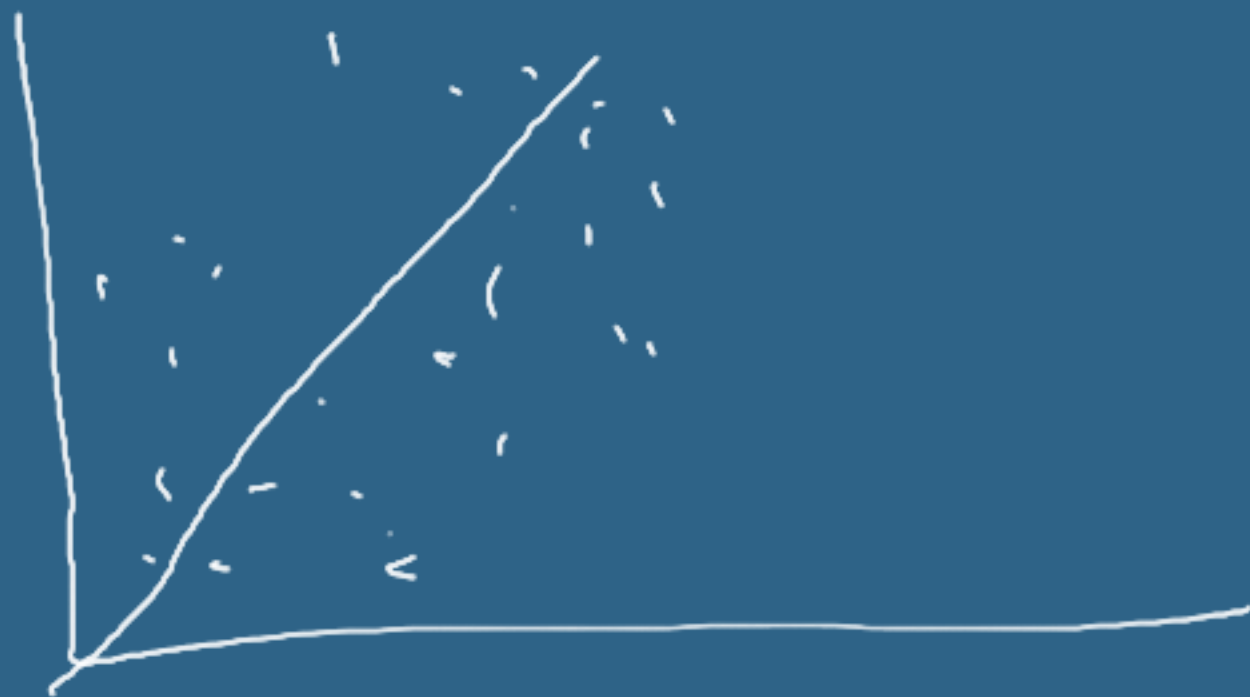


Polynomial

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



Underfitting Vs Overfitting



Evaluation metrics (Regression)

$R^2 \Rightarrow \%$ (0 - 1)

MSE \Rightarrow Value

Scaling

Min Max scaler
(0 - 1)

$$X_{\text{scaled}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$$

Standard scaler

$$X_{\text{scaled}} = \frac{X_i - X_{\text{mean}}}{\text{std}}$$

Robust scaler

$$X_{\text{scaled}} = \frac{X_i - X_{\text{median}}}{X_{Q_3} - X_{Q_1}}$$

↓

Not Robust to outliers

Robust to outliers

Gender $\begin{cases} \text{male} \\ \text{female} \end{cases}$

One hot Encoder

\rightarrow ~~X~~
Encoder fn

Gender \Rightarrow

male

female

Gender_male
1
0

0
1