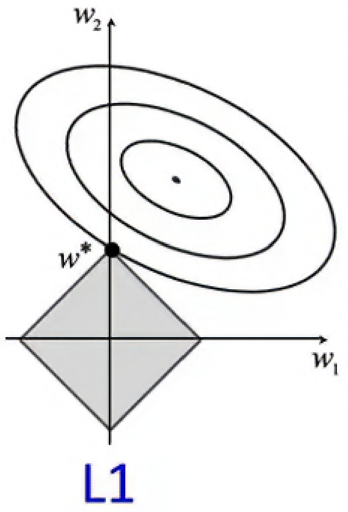
Lasso Regression

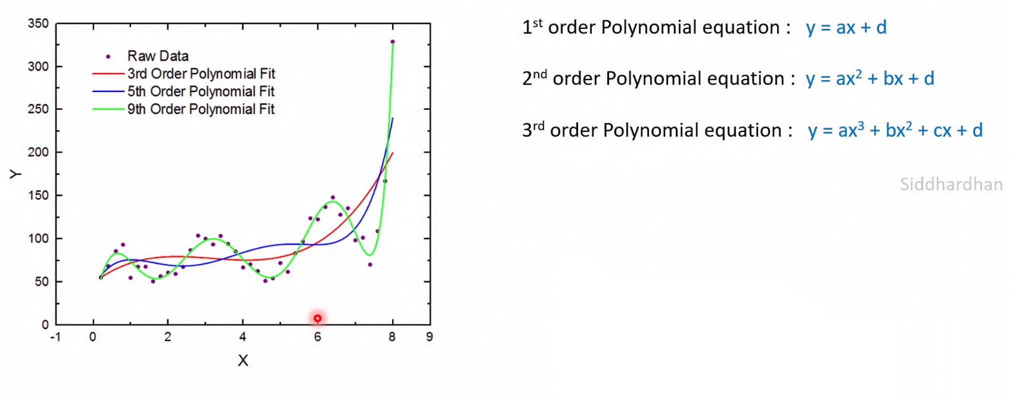
Created	@August 7, 2025
Edited	@August 8, 2025 5:11 AM
Archive	☐

It is built on top of linear regression model

1. Supervised Learning Model
2. Regression model
3. Least Absolute Shrinkage and Selection Operator(LASSO)
4. Implements regularization (L1) to avoid overfitting



Polynomial Equations:



X is independent variable

the green line is overfitted and regularization is required to avoid overfitting.

Inference: As the complexity of model increases, it tends to overfit the data.

Multiple Linear Regression:

It is a model for predicting the value of one dependent variable based on two or more independent variable.

Regularization:

Regularization is used to reduce the overfitting of the model by adding a penalty term(lambda) to the model, Lasso regression uses the L1 regularization technique.

The penalty term reduces the value of the coefficients or eliminate few coefficients, so that the model has fewer coefficients. As a result, overfitting can be avoided.

The process is called SHRINKAGE.

Math behind LASSO regression:

Cost Function for Lasso Regression :

$$J = \frac{1}{m}\left[\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} w_j\right]$$

m --> Total number of Data Points

n --> Total number of input features

$y^{(i)}$ --> True Value

$\hat{y}^{(i)}$ --> Predicted Value

λ --> Penalty Term

w --> Parameter of the model

Automatically can determine which features are important for the target variable and assigns weight value to the feature.

Gradient Descent for LASSO Regression:

$$w_2 = w_1 - L * \frac{dJ}{dw}$$

$$b_2 = b_1 - L * \frac{dJ}{db}$$

Gradients for Lasso Regularization

if (w_j > 0) :	***else (w_j ≤ 0) :***
$\frac{dJ}{dw} = \frac{-2}{m} \left[\left[\sum_{i=1}^{m} x_j \cdot \left(y^{(i)} - \hat{y}^{(i)} \right) \right] + \lambda \right]$	$\frac{dJ}{dw} = \frac{-2}{m} \left[\left[\sum_{i=1}^{m} x_j \cdot \left(y^{(i)} - \hat{y}^{(i)} \right) \right] - \lambda \right]$

Extra Penalty term to reduce overfitting problem.

$$\frac{dJ}{db} = \frac{-2}{m} \left[\sum_{i=1}^{m} \left(y^{(i)} - \hat{y}^{(i)} \right) \right]$$