ML Labs Pvt Ltd

RIDGE, LASSO, ELASTICNET REGRESSIONS

What are Ridge, Lasso, Elastic net regressions?

ML Labs Pvt Ltd

- **Ridge, Lasso, Elastic net Regressions** are Regularization techniques that are used for solving overfitting problem.

- These techniques **penalize the magnitude of coefficients** of features along with **minimizing the error** between predicted and actual observations.

Regularization:

Overfitting: ML model performing well on training data but poorly on validation (test) data.

Regularization helps to solve overfitting problem.

Regularization solves this by adding a penalty term to the objective function and controls the model complexity using that penalty term.

Regularization:

Regularization is generally useful in the following situations:

1. Large number of variables/features

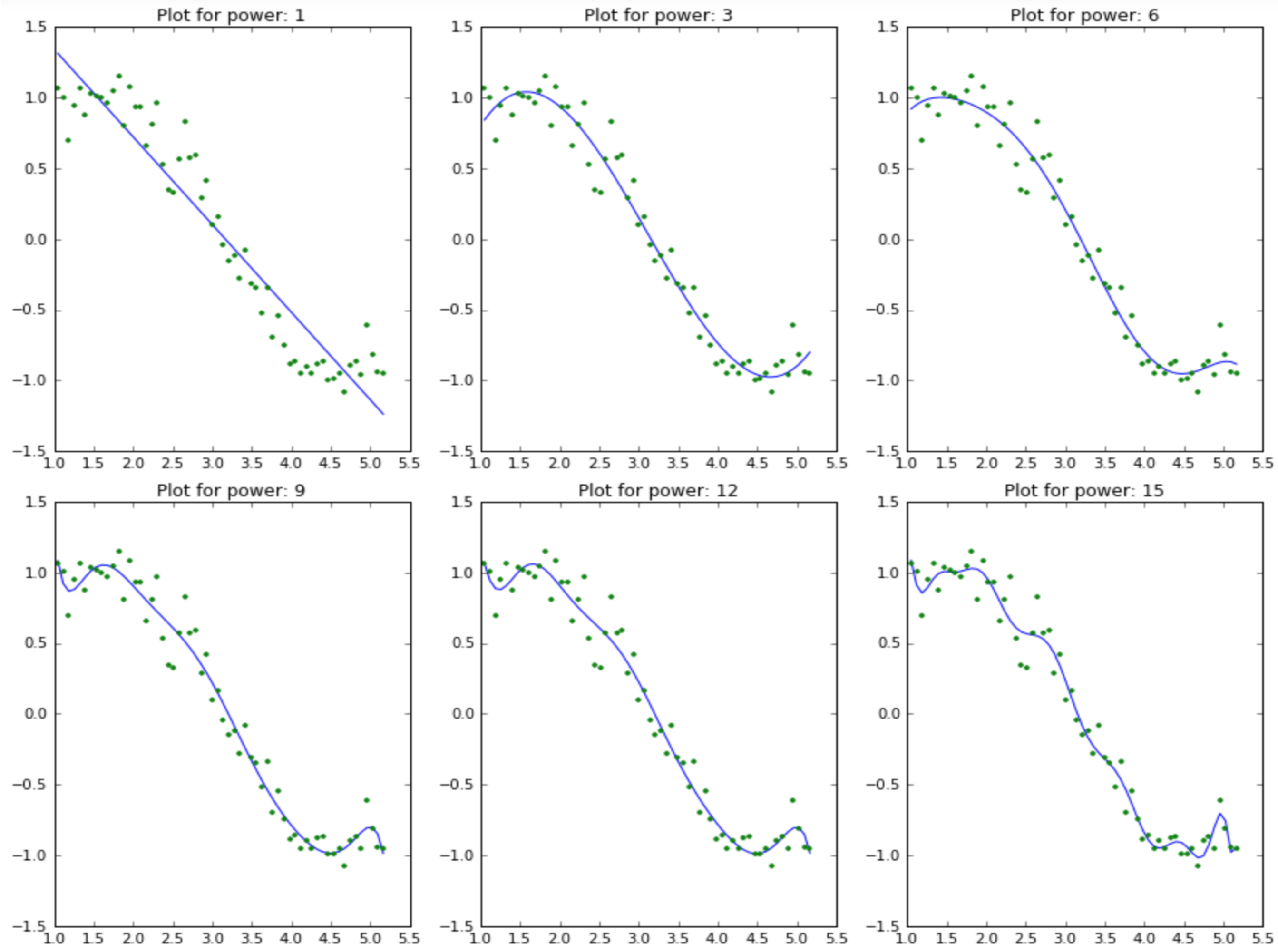
2. Low ratio of number of observations to number of variables/features

3. High Multi-Collinearity

Model Complexity Vs Overfitting

ML Labs Pvt Ltd

- With increase in complexity, models tend to fit even smaller deviations in the training data set.
- This leads to overfitting.
- The **number of coefficients increase exponentially with increase in model complexity**



Ridge Regression:

ML Labs Pvt Ltd

Ridge regression imposes an additional shrinkage penalty to the ordinary least squares loss function to limit its squared *L2* norm:

$$L(\bar{w}) = \|X\bar{w} - \bar{y}\|_2^2 + \alpha\|\bar{w}\|_2^2$$

Ridge Regression: Performs **L2 regularization**, i.e., adds penalty equivalent to **square of the magnitude** of coefficients

Minimization objective = Least Squares Objective + α * (sum of square of coefficients)

Ridge Regression:

α can take various values:

α = 0:

The objective becomes same as simple linear regression.

We'll get the same coefficients as simple linear regression.

Ridge Regression:

$\alpha = \infty$:

The coefficients will be zero, since infinite weightage on square of coefficients, anything other than zero will make the objective infinite.

Ridge Regression:

ML Labs Pvt Ltd

$0 < \alpha < \infty$:

The magnitude of α will decide the weightage given to different parts of objective.

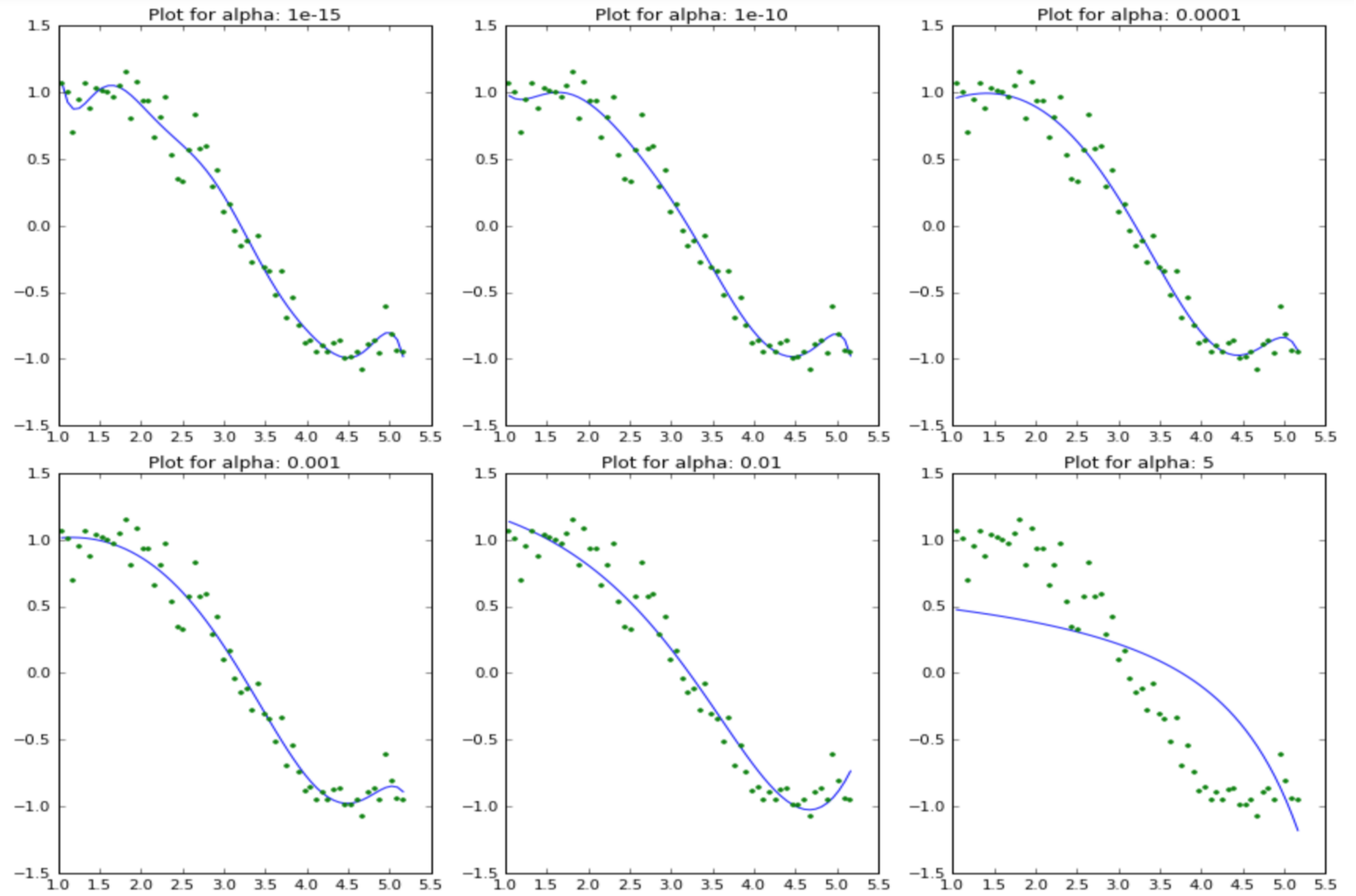
The coefficients will be somewhere between 0 and that of simple linear regression.

Alpha Vs Model Complexity:

ML Labs Pvt Ltd

The model complexity reduces as the value of alpha increases

High alpha values can lead to significant underfitting.



Lasso Regression:

Lasso Regression: Performs L1 regularization, i.e., adds penalty equivalent to **absolute value of the magnitude** of coefficients.

Minimization objective = LS Obj + α * (sum of absolute value of coefficients)

A Lasso regressor imposes a penalty on the *L1* norm of *w* to determine a potentially higher number of null coefficients:

$$L(\bar{w}) = \frac{1}{2n}\|X\bar{w} - \bar{y}\|_2^2 + \alpha\|\bar{w}\|_1$$

Lasso Regression:

ML Labs Pvt Ltd

Here, α (alpha) is same as that of ridge and provides a trade-off between balancing RSS and magnitude of coefficients.

$\alpha = 0$: Same coefficients as simple linear regression

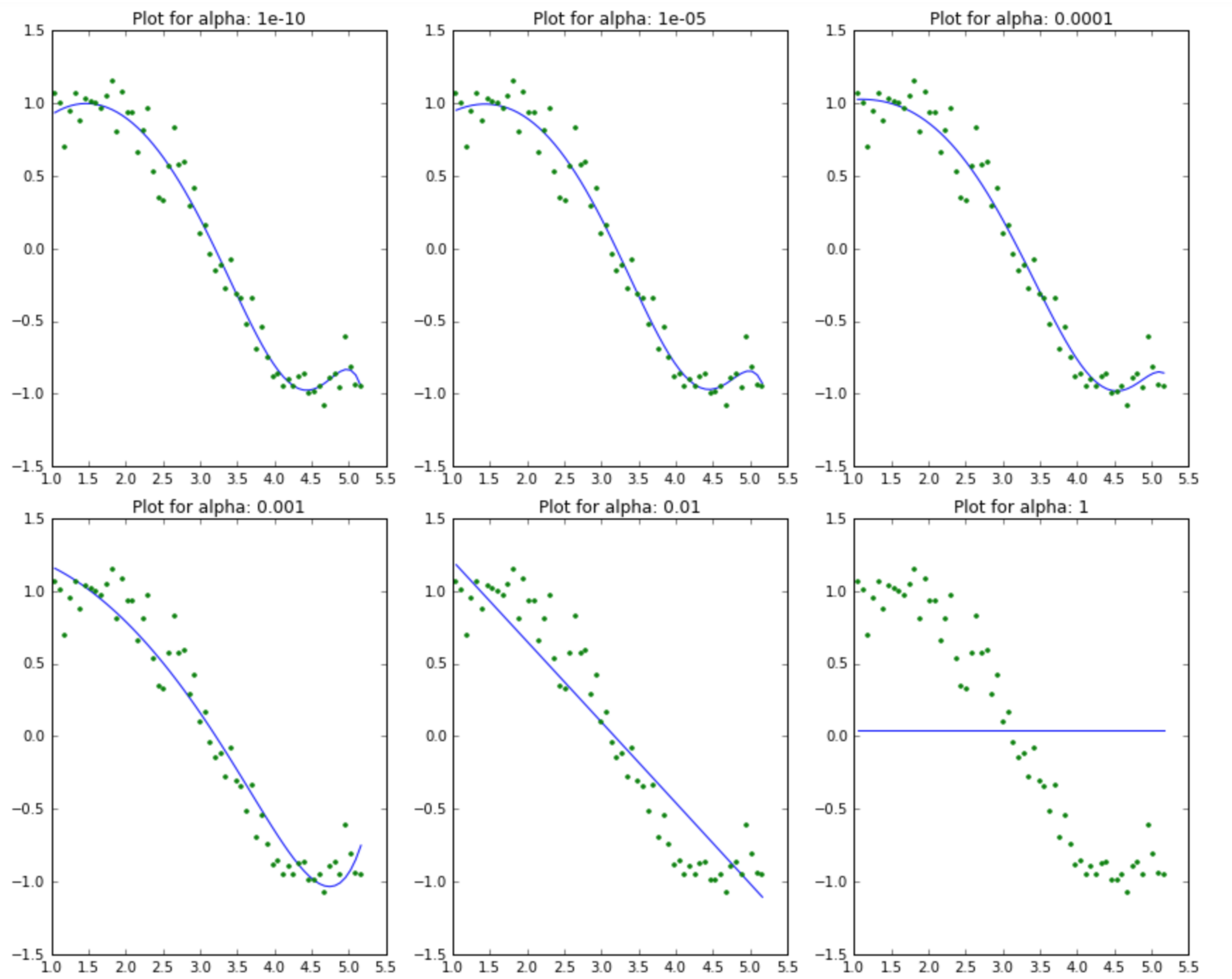
$\alpha = \infty$: All coefficients zero

$0 < \alpha < \infty$: coefficients between 0 and that of simple linear regression

Lasso Regression:

ML Labs Pvt Ltd

The model **complexity reduces** as the value of **alpha increases**



Ridge & Lasso

ML Labs Pvt Ltd

At higher alphas, RSS will be higher as well.

For the same values of alpha, the coefficients of lasso regression are much smaller as compared to that of ridge regression .

For the same alpha, lasso has higher RSS (poorer fit) as compared to ridge regression.

Many of the lasso regression coefficients are zero even for very small values of alpha.

ElasticNet

ML Labs Pvt Ltd

- ElasticNet is a Regularization technique that **combines both Lasso and Ridge into a single model with two penalty factors: one proportional to *L1* norm and the other to *L2* norm**.

- ElasticNet model will be sparse like a pure Lasso, but with the same regularization ability as provided by Ridge. The resulting loss function is:

$$L(\bar{w}) = \frac{1}{2n}\|X\bar{w} - \bar{y}\|_2^2 + \alpha\beta\|\bar{w}\|_1 + \frac{\alpha(1-\beta)}{2}\|\bar{w}\|_2^2$$

Summary

L1 Regularization aka Lasso Regularization

ML Labs Pvt Ltd

–Adds regularization terms in the model that are function of **absolute value** of the coefficients of parameters.

—The coefficients of the parameters can be driven to zero as well during the regularization process. Hence **this technique can be used for feature selection.**

—Not good for grouped selection for highly correlated features.

Summary

ML Labs Pvt Ltd

L2 Regularization aka Ridge Regularization

—This adds regularization terms in the model which are function of square of coefficients of parameters. Coefficient of parameters can approach to zero but never become zero.

—Good for multi-collinearity.

—Not Good for feature selection.

Summary

Elastic Net

—This adds regularization terms in the model which are combination of both L1 and L2 regularization.

—Trades bias for variance reduction

—better prediction accuracy