# Lasso and Ridge Regressions and Elastic Net

## 1. Lasso Regression

Lasso Regression, or Least Absolute Shrinkage and Selection Operator, is a type of linear regression that includes an L1 regularization term. This regularization term encourages the coefficients of less important features to be exactly zero, thus performing feature selection.

### Key Features:

• L1 regularization: Encourages sparsity in the model.  
• Feature selection: Automatically selects the most important features by shrinking less important ones to zero.

### Hyperparameters:

• \*\*Alpha (λ):\*\* Controls the strength of the L1 regularization. Higher values of alpha result in more coefficients being shrunk to zero.

### Steps:

1. \*\*Formulate Objective Function:\*\* Include the L1 regularization term in the loss function.  
2. \*\*Optimize:\*\* Use an optimization algorithm to minimize the loss function.  
3. \*\*Predict:\*\* Use the resulting model to make predictions on new data.

### Advantages:

• Performs feature selection and regularization simultaneously.  
• Can handle high-dimensional data well.

### Disadvantages:

• May not perform well when the number of observations is less than the number of features.  
• Can lead to bias in the estimates of the coefficients.

## 2. Ridge Regression

Ridge Regression, also known as Tikhonov regularization, is a type of linear regression that includes an L2 regularization term. This term helps to prevent overfitting by shrinking the coefficients of less important features.

### Key Features:

• L2 regularization: Penalizes the sum of the squared coefficients.  
• Reduces overfitting: By adding a penalty to the size of the coefficients.

### Hyperparameters:

• \*\*Alpha (λ):\*\* Controls the strength of the L2 regularization. Higher values of alpha result in more shrinkage.

### Steps:

1. \*\*Formulate Objective Function:\*\* Include the L2 regularization term in the loss function.  
2. \*\*Optimize:\*\* Use an optimization algorithm to minimize the loss function.  
3. \*\*Predict:\*\* Use the resulting model to make predictions on new data.

### Advantages:

• Reduces overfitting by shrinking coefficients.  
• Can handle multicollinearity in the data.

### Disadvantages:

• Does not perform feature selection; all features are retained in the model.

## 3. Elastic Net

Elastic Net is a regularized regression method that linearly combines L1 and L2 penalties of the Lasso and Ridge methods. It is useful when there are multiple features that are correlated with each other.

### Key Features:

• Combination of L1 and L2 regularization: Balances the benefits of both Lasso and Ridge.  
• Suitable for high-dimensional data: Performs well when there are many correlated features.

### Hyperparameters:

• \*\*Alpha (λ):\*\* Controls the overall strength of the regularization.  
• \*\*L1 ratio:\*\* Defines the mix between L1 and L2 regularization. A value of 0 corresponds to Ridge, a value of 1 corresponds to Lasso, and values between 0 and 1 correspond to Elastic Net.

### Steps:

1. \*\*Formulate Objective Function:\*\* Include both L1 and L2 regularization terms in the loss function.  
2. \*\*Optimize:\*\* Use an optimization algorithm to minimize the loss function.  
3. \*\*Predict:\*\* Use the resulting model to make predictions on new data.

### Advantages:

• Combines the strengths of both Lasso and Ridge regressions.  
• Can perform well when dealing with correlated features.

### Disadvantages:

• More complex to tune due to the presence of two regularization parameters.  
• May require more computational resources compared to Lasso or Ridge alone.

## Conclusion

Lasso, Ridge, and Elastic Net are powerful regularization techniques that help to prevent overfitting and manage high-dimensional data. Lasso performs feature selection by shrinking some coefficients to zero, Ridge shrinks all coefficients but retains them, and Elastic Net combines the benefits of both. Proper tuning of the hyperparameters is crucial for these methods to achieve optimal performance.