Regularization

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

Recap

By the End of this Session, You Will:

- Understand the concept of overfitting
- Explore the concept of bias-variance trade-off in model performance
- Understand Regularization and differentiate between various types of regularization methods
- Learn about LASSO Regression
- Understand the loss function used in LASSO Regression

What’s in It for Me?

Understand the concept of overfitting and identify the reasons for overfitting in machine learning models.

Learn how to improve model generalization by avoiding overfitting and enhance decision-making through a deeper grasp of model biases and variances.

Define regularization in the context of machine learning and explore the motivations behind using regularization techniques.

Learn how LASSO Regression differs from other regression techniques and discover how LASSO can be employed as a feature selection method.

Poll Time

Q. In a linear regression equation, what does the slope (coefficients) represent?

a. The intercept of the regression line

b. The variance of the dependent variable

c. The change in the dependent variable for a one-unit change in the independent variable

d. The correlation between the variables

Poll Time

Q. In a linear regression equation, what does the slope (coefficients) represent?

a. The intercept of the regression line

b. The variance of the dependent variable

c. **The change in the dependent variable for a one-unit change in the independent variable**

d. The correlation between the variables

Overfitting

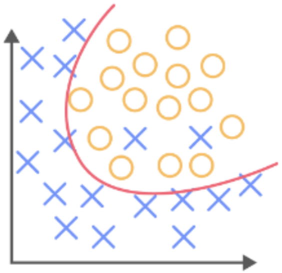
What is Overfitting?

Definition – Model learns training data too well to the point that it perform poorly on unseen or new data. Model captures noise in training data rather than underlying patterns.

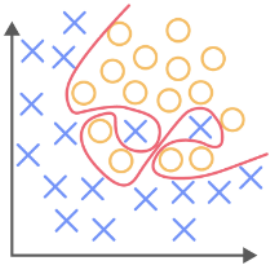
Test data performance – When you use an overfit model to make predictions on new data, it tends to perform poorly because it has essentially memorized training data.

High number of parameters – Complex models have a high number of parameters, allowing them to fit the training data very closely. In case of overfitting, the model becomes too "sensitive" to training data.

Low training error – Model fits training data perfectly, achieving a very low training error. However, it doesn't perform well on new, unseen data.



Appropriate fitting



Overfitting

Reasons for Overfitting

Overly complex model architectures – Models with many parameters fit training data very closely, including its noise.

Small amount of training data – Relative to the complexity of the model. model doesn't have enough information to generalize. Hence, models memorize the individual data points.

Noise in the data - Noise refers to random variations or errors (measurement errors, data collection issues) in data. Model may mistakenly try to fit this noise, leading to overfitting.

Irrelevant Features – Including irrelevant features or variables in the model. Irrelevant features often contain noise or random variations that don't have any meaningful relationship with the target variable.

Bias

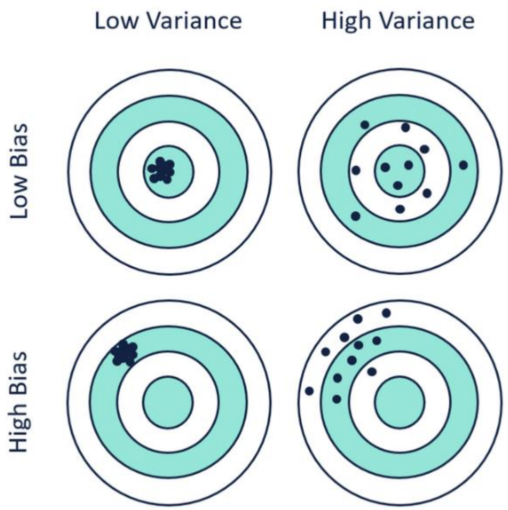
Bias represents the model's ability to capture true underlying patterns in the data.

High bias models are overly simplistic and make strong assumptions about the data, leading to underfitting (too simple model unable to capture patterns and performs poorly on both training and test data).

The model should be low biased so that training and testing errors will be low.

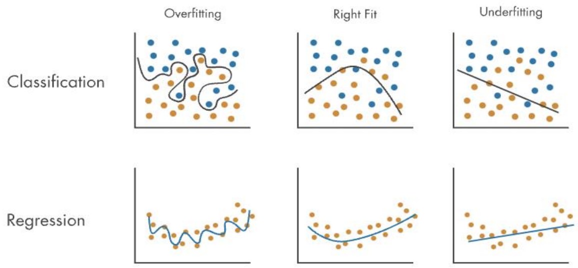
How to Identify High Bias:

- High training error (poor fit to training data).
- High test error (poor generalization to new data).
- Visual inspection may reveal that the model does not capture data's complexity.



Underfitting vs. Overfitting

Underfitting	Overfitting
It doesn't fit training data well and performs poorly on both training and test (or validation) data.	Model fits training data perfectly, including its noise and random fluctuations. Fails to generalize to new, unseen data.
The model has high bias and low variance.	The model has low bias and high variance.
Too few parameters to capture complexity of data.	Too many parameters or are overly flexible.
Not including enough relevant features.	Lack of sufficient training data, especially in high-dimensional spaces.
High training error and high test error.	Low training error but high test error.
Visual inspection : Model's predictions do not fit data points well.	Visual inspection : Model fits training data well but fails to generalize.



Variance

Variance measures model's sensitivity to small fluctuations in training data. It represents model's ability to adapt to different variations in training data.

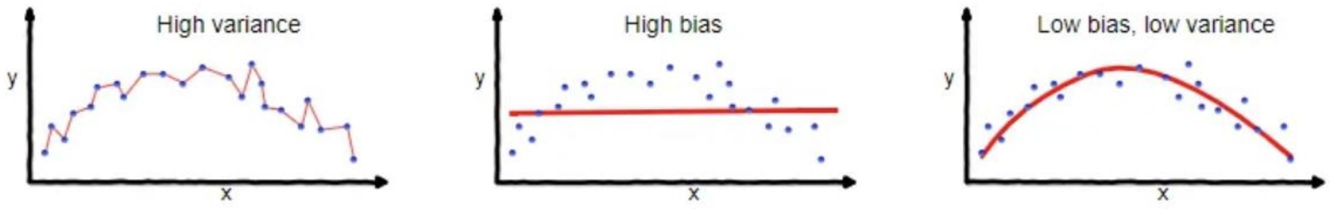
1. High variance models are overly complex and flexible, leading to overfitting.
2. High variance models capture not only underlying patterns but also the noise.

Causes of High Variance:

1. Using a complex model with many parameters.
2. Having insufficient training data, especially in high-dimensional spaces.

How to Identify High Variance:

- Low training error (good fit to training data).
- High test error (poor generalization to new data).
- Visual inspection : Model fits training data too closely



Demo : Overfitting

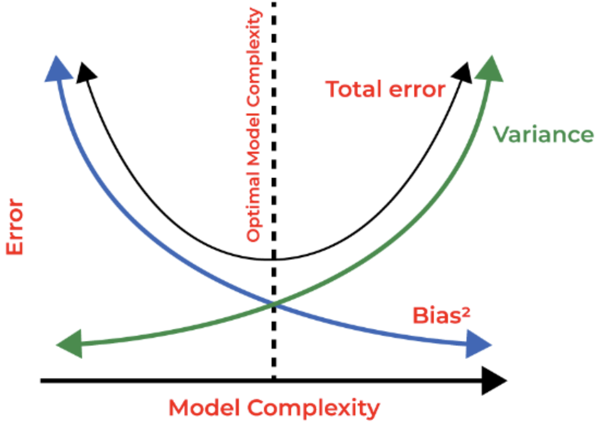
Bias-Variance Trade-off

- Increasing model complexity (e.g., using more features or a more complex algorithm) tends to decrease bias but increases variance.
- Conversely, reducing model complexity (e.g., using fewer features or a simpler algorithm) tends to increase bias but decrease variance.

Balance between bias and variance to achieve a model that generalizes well to new, unseen data while still capturing essential patterns in training data.

- More complex models have the potential to fit training data closely, but they are more prone to overfitting.
- Simpler models are more robust but potentially less accurate.

A good model minimizes both bias and variance, achieving low training error and low test error. To find balance, techniques like regularization (to reduce variance) and feature engineering (to reduce bias), as well as model selection and hyperparameter tuning.



Demo : Bias-Variance Trade-off

Poll Time

Q. Which of the following statements is true regarding the bias-variance trade-off?

a. Models with high bias and low variance typically underfit the data

b. Models with low bias and high variance tend to generalize well to new data

c. The goal is to find a balance between bias and variance to minimize test error

d. Increasing bias always leads to better model performance

Poll Time

Q. Which of the following statements is true regarding the bias-variance trade-off?

a. Models with high bias and low variance typically underfit the data

b. Models with low bias and high variance tend to generalize well to new data

c. **The goal is to find a balance between bias and variance to minimize test error**

d. Increasing bias always leads to better model performance

Regularization

Introduction to Regularization

- Regularization is used to prevent overfitting and improve the generalization performance of models.
- It involves adding a penalty term to model's loss function to discourage extreme or overly complex parameter values.
- Regularization plays a crucial role in finding the right balance between bias and variance in a model.

- Regularization adds a penalty for complexity, encouraging the model to be less complex and more focused on capturing essential patterns in the data.
- By adjusting the strength of the regularization penalty, you can control the balance between bias and variance in the model.
- A stronger regularization penalty results in simpler models with lower variance but potentially higher bias.
- Regularization allows practitioners to strike a balance between fitting the training data and building models that generalize effectively to new data.

Why is Regularization Needed?

- Overfit models have excessively complex parameter values that lead to poor generalization performance on unseen data.

- Regularization helps control this complexity and encourages models to generalize better.

- In high-dimensional datasets with many features, models can become more prone to overfitting because they have more opportunities to capture noise.

- Regularization is effective at feature selection by automatically identifying and using only the most relevant features, reducing the risk of overfitting in high-dimensional spaces.

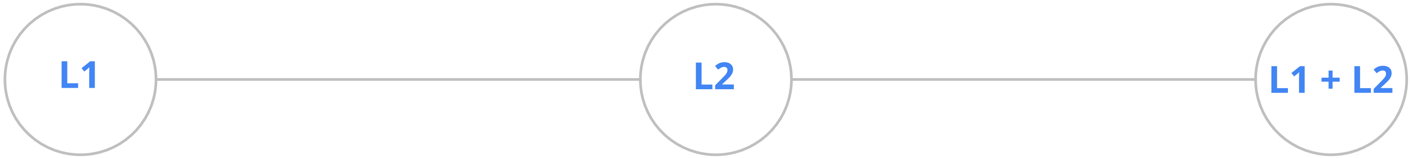
- In datasets where features are highly correlated (multicollinearity), Regularization helps by discouraging extreme coefficient values. This leads to more stable parameter estimates.

- Regularization contributes to the robustness of machine learning models by discouraging extreme or irregular parameter values, thus making models less sensitive to minor changes in training data.

Different Types of Regularization

Each type of regularization has its unique characteristics and application areas.

Here are the details of some common types of regularization of linear models:



L1 regularization, also known as Lasso (Least Absolute Shrinkage and Selection Operator), adds the absolute values of the model's coefficients as a penalty term to the loss function.

L2 regularization, also known as Ridge regularization, adds the squared values of the model's coefficients as a penalty term to the loss function.

Elastic Net regularization combines both L1 and L2 regularization by adding both the absolute values (L1) and squared values (L2) of the model's coefficients as penalty terms to the loss function.

Pop Quiz

Q. In which situation would you typically consider using regularization?

a. When the model has a low bias
b. When the model has a high variance
c. When the model fits the training data perfectly
d. When the model has many features

Pop Quiz

Q. In which situation would you typically consider using regularization?

a. When the model has a low bias

b. When the model has a high variance

c. When the model fits the training data perfectly

d. When the model has many features


Q&A

Lasso Regression

What is LASSO Regression?

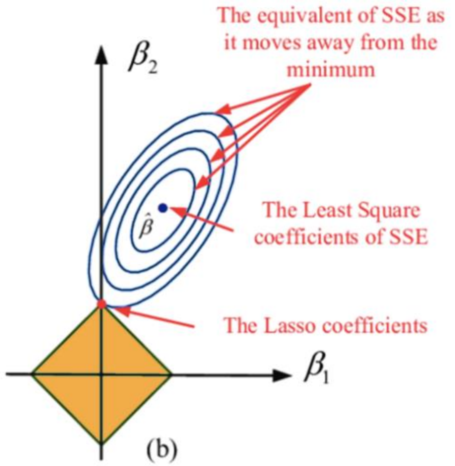
- LASSO Regression, which stands for "Least Absolute Shrinkage and Selection Operator," is a type of regularization technique used in linear regression and related models.
- LASSO is designed to address issues like multicollinearity and overfitting by adding a penalty term to the linear regression's cost function.
- LASSO Regression aims to improve the accuracy and interpretability of linear regression models while handling situations where there are many predictor variables (features), and some of them may be irrelevant or highly correlated.

L1 Penalty Term

- The key feature of LASSO is the addition of an L1 regularization penalty term to the linear regression's cost function.
- This penalty is represented as $\lambda * \sum|\beta|$, where β represents the coefficients of the linear regression model, and λ is the regularization strength parameter.
- The L1 penalty term encourages sparsity in coefficient values, effectively driving some coefficients to become exactly zero. This means that LASSO can perform automatic feature selection by excluding irrelevant variables from the model.

How to Shrink the Beta Coefficients?

- The objective of linear regression is to minimize the residual sum of squares (RSS), which corresponds to finding the coefficients that yield the smallest error.
- L1 regularization introduces a constraint in form of a diamond-shaped region (L1-norm constraint) in the coefficient space. The coefficients must lie within this region while minimizing RSS.
- L1 constraint often results in the coefficients being driven to the corners (vertices) of the diamond-shaped region.
- Since each corner corresponds to some coefficients being exactly zero, LASSO has the effect of zeroing out certain coefficients.
- The decision of which coefficients to shrink or set to zero depends on the value of λ, the regularization strength parameter. A larger λ will result in stronger shrinkage and more coefficients being set to zero.
- LASSO's shrinkage of coefficients represents a trade-off between model fit and model simplicity. As you increase λ, the model becomes, and its fit to training data decreases. This helps control model complexity and prevent overfitting.



Benefits of Shrinking the Beta Coefficients

Shrinking the beta coefficients in LASSO Regression has several benefits:

- **Feature Selection:** LASSO automatically selects the most important features by setting others to zero, simplifying the model and improving interpretability.
- **Reducing Overfitting:** By discouraging extreme coefficient values, LASSO helps prevent overfitting, making the model more robust and better at generalizing to unseen data.
- **Handling Multicollinearity:** LASSO can handle multicollinearity (correlation between predictor variables) by selecting one variable from a group of highly correlated variables.

After performing LASSO Regression, the coefficients that are not set to zero provide insights into the variables that have the most significant impact on the response variable. These nonzero coefficients represent the predictor variables that are selected by the model.

LOSS Function in LASSO

The loss function, also known as the Objective function, quantifies how well the model's predictions align with actual target values. The loss function includes two main components: the traditional least squares (LS) loss and the L1 regularization penalty.

Least Squares (LS) Loss Component:

LS Loss = $\Sigma(y_i - \hat{y}_i)^2$ for i = 1 to n

It quantifies the difference between the model's predictions and actual target values. The goal is to minimize this component to achieve a good fit to the training data. LS loss component aims to minimize RSS.

L1 Regularization Penalty Component:

L1 Penalty = $\lambda * \Sigma|\beta|$ for all coefficients β
λ represents the regularization strength parameter

Total Loss = LS Loss + L1 Penalty = $\Sigma(y_i - \hat{y}_i)^2 + \lambda * \Sigma|\beta|$ for i = 1 to n

The goal in LASSO is to minimize this combined loss function. The first part (LS Loss) ensures that the model fits the training data well, while the second part (L1 Penalty) encourages simplicity and sparsity in the model by shrinking some coefficients toward zero.

Demo : LASSO Regression

LASSO as a Feature Selection Method

- It is particularly valuable when you have a large number of predictor variables (features) and want to identify the most relevant ones while excluding irrelevant or redundant features.
- LASSO encourages sparsity in the coefficient values, meaning it drives some coefficients to become exactly zero. As a result, certain predictor variables are effectively excluded from the model.
- The algorithm identifies which predictor variables have a substantial impact on the response variable and keeps them in the model while setting the coefficients of less influential variables to zero.
- By examining the magnitude and sign of the non-zero coefficients, you can understand how each selected feature contributes to the response variable.
- A larger λ results in stronger shrinkage and more coefficients being set to zero, which means more features are excluded from the model. A smaller λ allows more coefficients to remain non-zero.
- LASSO can effectively handle multicollinearity (correlation between features) by selecting one variable from a group of correlated variables.

Demo : Feature Selection Using LASSO Regression

Pop Quiz

Q. In Lasso Regression, what type of penalty term is added to the linear regression cost function?

a. No penalty term is added

b. A penalty term proportional to the absolute values of the coefficients

c. A penalty term proportional to the squared values of the coefficients

d. A penalty term proportional to the square root of the coefficients

Pop Quiz

Q. In Lasso Regression, what type of penalty term is added to the linear regression cost function?

a. No penalty term is added

b. A penalty term proportional to the absolute values of the coefficients

c. A penalty term proportional to the squared values of the coefficients

d. A penalty term proportional to the square root of the coefficients


Q&A

Activity 1

Pre-requisites:

Participants should have a basic understanding of probability and the concept of bias and variance.

Scenario:

Imagine you are flipping the coin repeatedly and recording the outcomes. You want to understand the tradeoff between bias and variance when estimating the average result.

Data:

Flip the coin 20 times and record the outcomes. Use "H" for heads and "T" for tails. For example:
Outcomes: H, T, H, H, T, H, T, H, T, T, H, T, T, H, H, H, T, T, T, H

Activity 1

Expected Outcome:

Gain an understanding of the bias-variance tradeoff in a simple scenario involving coin flipping.

Steps:

1. Use each individual flip's outcome as an estimate of the probability and calculate the estimated probability (Estimator 1).

2. Average outcomes of multiple flips to estimate the probability and calculate the estimated probability (Estimator 2).

3. Compare the bias and variance of Estimator 1 and Estimator 2 and discuss their tradeoffs.

Activity 2

Pre-requisites:

Participants should have a basic understanding of linear regression and LASSO regression.

Scenario:

You are working with a dataset of home prices and want to build a regression model to predict home prices based on two features: square footage and the number of bedrooms. You've heard about Lasso Regression and want to explore how it affects the model's coefficients.

Data: Refer to the table on the right side

Square Footage (sq. ft.)	Bedrooms	Home Price (USD)
1500	3	200,000
1800	2	220,000
2000	4	250,000
1400	2	180,000
1600	3	210,000

Activity 2

Expected Outcome:

Understand the concept of Lasso Regression as a regularization technique. Recognize how Lasso Regression affects model coefficients. Apply Lasso Regression to a simple dataset with two features.

Steps:

1. Build an ordinary linear regression model to predict home prices using the two features: Square Footage and Bedrooms.

2. Apply Lasso Regression to the dataset with a specified regularization parameter (alpha).

3. Calculate and discuss the coefficients obtained after Lasso Regression.

4. Highlight how Lasso Regression can shrink coefficients towards zero, effectively performing feature selection.

Summary

- Overfitting occurs when a model learns the training data too well but struggles to generalize to new, unseen data.
- Regularization is a technique to prevent overfitting by adding a penalty term to the model's cost function.
- LASSO Regression is a specific regularization method that acts as both a predictive model and a feature selection technique.
- LASSO achieves feature selection by shrinking certain coefficient values to zero, effectively excluding those features from the model.

Next Session:
Ridge and Elastic Net Regression
Regularization – Case Study

THANK YOU!

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad

Ridge and Elastic Net Regression Case Study on Regularization

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of this Session, You Will:

- Learn the definition and principles of Ridge Regression
- Understand Elastic Net Regression as a combination of L1 and L2 regularization
- Explore real world data and perform regularization on it
- Conduct extensive exploratory data analysis before creating the model and write its conclusions

Recap

Poll Time

Q. When the value of the regularization parameter (λ) in Lasso Regression is increased, what happens to the coefficients of less important features?

a. They remain unchanged

b. They increase in magnitude

c. They decrease in magnitude

d. They are set to zero

Poll Time

Q. When the value of the regularization parameter (λ) in Lasso Regression is increased, what happens to the coefficients of less important features?

a. They remain unchanged

b. They increase in magnitude

c. They decrease in magnitude

d. They are set to zero

Ridge Regression

What is Ridge Regression?

Ridge regression is useful in addressing the problem of multicollinearity and overfitting in linear regression models.

When two or more independent variables (features) are highly correlated (multicollinearity), it can be challenging to determine their individual contributions to the target variable. Ridge Regression addresses this issue by stabilizing coefficients.

Ridge Regression introduces a regularization term (L2 penalty) to the linear regression cost function. The cost function for Ridge Regression is as follows:
Cost = $(1/2m) * \Sigma(y_i - \hat{y}_i)^2 + \lambda * \Sigma\beta^2$

- The first part is the ordinary least squares (OLS) term, which minimizes the error between predicted ($\hat{y}$) and actual (y) values.
- The second part is the regularization term. λ (lambda) is the regularization parameter, which controls the strength of the regularization. Larger values of λ lead to stronger regularization.

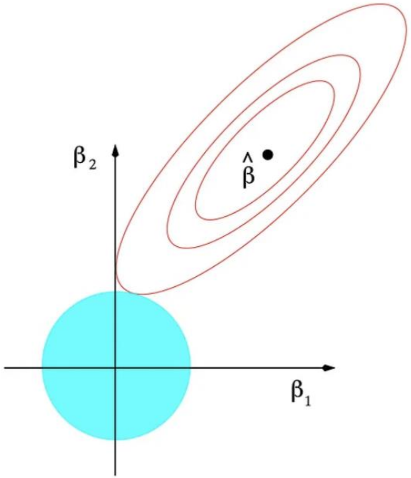
Ridge Regression does not eliminate features entirely but shrinks their coefficients driving them towards zero. This means that all features contribute to the prediction, but some have a more minor impact due to their smaller coefficients.

The choice of λ is crucial in Ridge Regression. A small λ may not provide enough regularization, while a large λ may excessively penalize coefficients.

Demo : Ridge Regression

How to Shrink Beta Coefficients Using Ridge Regression?

- Here's how Ridge Regression effectively shrinks the beta coefficients:
- Cost = $(1/2m) * \Sigma(y_i - \hat{y}_i)^2 + \lambda * \Sigma\beta^2$
- The second part is regularization term. λ (lambda) is regularization parameter, and $\Sigma\beta^2$ represents sum of squared beta coefficients for all features except intercept term.
- In Ridge Regression, a different form of constraint is introduced in the coefficient space, which takes the shape of a circle (L2-norm constraint) instead of a diamond.
- The coefficients of the model must lie within this circular region while minimizing the Residual Sum of Squares (RSS).
- Due to the circular constraint, Ridge Regression tends to shrink the coefficients toward the center of the circle. This means that all coefficients are penalized, but none are driven to absolute zero.
- The degree of shrinkage is controlled by the regularization parameter λ. Higher values of λ lead to stronger regularization, causing the coefficients to be shrunk more aggressively.



Demo : Handling Multicollinearity using Ridge Regression

Pop Quiz

Q. What type of penalty term is added to the linear regression cost function in Ridge Regression?

a. No penalty term is added
b. A penalty term proportional to the absolute values of the coefficients
c. A penalty term proportional to the squared values of the coefficients
d. A penalty term proportional to the square root of the coefficients

Pop Quiz

Q. What type of penalty term is added to the linear regression cost function in Ridge Regression?

a. No penalty term is added

b. A penalty term proportional to the absolute values of the coefficients

c. A penalty term proportional to the squared values of the coefficients

d. A penalty term proportional to the square root of the coefficients

Elastic Net Regression

What is Elastic Net Regression?

It combines characteristics of both Lasso Regression (L1 regularization) and Ridge Regression (L2 regularization). Elastic Net addresses the limitations of each of these techniques while offering a flexible way to handle multicollinearity, overfitting, and feature selection.

Elastic Net combines both L1 (Lasso) and L2 (Ridge) regularization terms in linear regression cost function. The cost function for Elastic Net is defined as:
$\text{Cost} = (1/2m) * \Sigma(y_i - \hat{y}_i)^2 + \lambda_1 * \Sigma|\beta| + \lambda_2 * \Sigma\beta^2$

- The first part is the ordinary least squares (OLS) term
- The second part represents L1 regularization, with λ_1 controlling the strength of L1 penalty term.
- The third part represents L2 regularization, with λ_2 controlling the strength of L2 penalty term.

Elastic Net simultaneously encourages feature selection (by setting some coefficients to zero) and penalizes the magnitude of the coefficients (to prevent them from becoming too large).

Elastic Net provides a balance between the strengths of Lasso (feature selection) and Ridge (coefficients shrinkage) Regression.

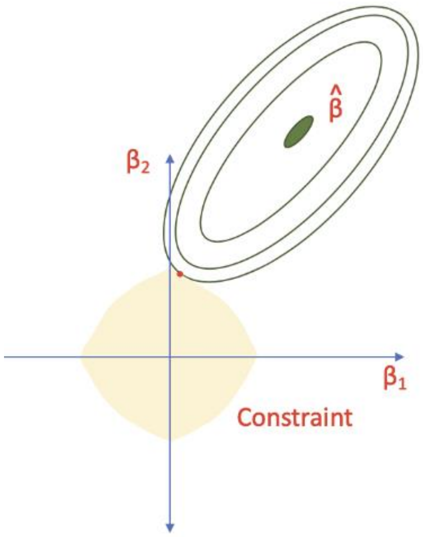
How to Shrink Beta Coefficients Using Elastic Net?

- In Elastic Net Regression, the primary goal is to shrink beta coefficients (regression coefficients) associated with features to prevent overfitting and address multicollinearity.

- The L1 regularization term encourages feature selection by setting some feature coefficients (beta values) to exactly zero. This means that some features are effectively eliminated from the model as their coefficients become zero.

- The L2 regularization term encourages coefficient shrinkage by penalizing the magnitude of the coefficients. This prevents the coefficients from becoming too large, which helps prevent overfitting.

- Higher values of λ_1 encourage stronger L1 regularization, which results in more feature selection, while higher values of λ_2 encourage stronger L2 regularization, which leads to more coefficient shrinkage.



Demo : Elastic Net Regression

Pop Quiz

Q. How does L1 (Lasso) regularization in Elastic Net affect feature coefficients?

a. It encourages feature selection by setting some feature coefficients to zero

b. It increases the magnitude of all feature coefficients

c. It has no impact on feature coefficients

d. It eliminates all features except one

Pop Quiz

Q. How does L1 (Lasso) regularization in Elastic Net affect feature coefficients?

a. **It encourages feature selection by setting some feature coefficients to zero**

b. It increases the magnitude of all feature coefficients

c. It has no impact on feature coefficients

d. It eliminates all features except one

Demo : Applying and Comparing All 3 Regularization Techniques


Q&A
?
?
?
?
?
?
?
?

Case Study on Regularization

Case Study – Problem Statement

Problem Statement

- One of the real estate organization is looking to predict the house price based on the various features like avg. Rooms per building, residential zone, distance from highways, etc.

- They are having the online business and wanted to implement the same so that if any new client or individual comes into their website, they will be asked to input their preferences using which the organization wants to create a model.

- To achieve this goal, Once a new customer comes in, the machine will be able to predict if the housing price. You must apply the regularization technique to solve this problem statement.

- The dataset comprises several columns, including lavg. Rooms per building, residential zone, distance from highways, etc. and price as target variable

- Based on these features, this company would like to predict housing price by using Linear Regression and use the concept of regularization.

- Also apply all the regularization techniques learnt during the session, i.e., Lasso (L1), Ridge (L2), and Elastic net.

Areas to Focus

Pre-processing the data

Outlier treatment,
missing value treatment,
train test split, etc.

Model Evaluation

MSE, MAE, RMSE,
R2 Score etc.

Apply Regularization

Lasso,
Ridge,
Elastic Net

Hands-on: Case Study Questions

Poll Time

Q. Which type of regularization is more likely to result in feature selection, where some features have zero coefficients?

a. L1 Regularization (LASSO)

b. L2 Regularization (Ridge)

c. Elastic Net Regularization

d. No regularization can achieve feature selection

Poll Time

Q. Which type of regularization is more likely to result in feature selection, where some features have zero coefficients?

a. **L1 Regularization (LASSO)**

b. L2 Regularization (Ridge)

c. Elastic Net Regularization

d. No regularization can achieve feature selection

Poll Time

Q. Which of the following statements is true regarding regularization?

a. Regularization always improves model accuracy on the training data

b. Regularization can help prevent both overfitting and underfitting

c. Regularization is only applicable to linear models

d. Regularization has no impact on the model's complexity

Poll Time

Q. Which of the following statements is true regarding regularization?

a. Regularization always improves model accuracy on the training data

b. Regularization can help prevent both overfitting and underfitting

c. Regularization is only applicable to linear models

d. Regularization has no impact on the model's complexity


Q&A

Activity 1

Pre-requisites:

Familiarity with python pandas library, linear regression concepts, and ridge regression as a regularization technique.

Scenario:

You are a data scientist working for a real estate agency. Your goal is to develop a predictive model for housing prices based on various features of houses. However, you've heard that there might be issues with multicollinearity in the dataset. To address this issue and improve the model's generalization, you decide to apply Ridge Regression.

Data:

```
import pandas as pd
df = pd.DataFrame({ "Size (in sq. ft.)": [1500, 2000, 1600, 1900, 2100, 1700],
"Distance to School (miles)": [0.5, 0.3, 0.4, 0.3, 0.6, 0.4],
"Price (in $1000s)": [300, 400, 350, 370, 410, 340] })
```

Activity 1

Expected Outcome:

Built a Ridge regression model using the dataset. Implement Ridge Regression using scikit-learn. Evaluate the model's performance using relevant metrics.

Steps:

1. Load and explore the dataset
2. Split the dataset into the feature and target variables.
3. Create a Ridge Regression model.
4. Evaluate the model's performance using metrics like Mean Squared Error (MSE) and R-squared.
5. Compare the results with a standard linear regression model.

Activity 2

Pre-requisites:

Familiarity with python pandas library, linear regression concepts and elastic net regression as a regularization technique.

Scenario:

You work for a car manufacturing company, and you want to develop a predictive model to estimate the fuel efficiency (miles per gallon, MPG) of cars based on two key features: engine displacement (in liters) and vehicle weight (in kilograms). However, you suspect that there may be multicollinearity issues in the dataset. To address this and build a robust model, you decide to apply Elastic Net Regression.

Data:

```
import pandas as pd
df = pd.DataFrame({'Engine_Displacement': [1.6, 2.0, 1.8, 2.2, 1.5, 2.4],
 'Vehicle_Weight': [1200, 1400, 1300, 1500, 1100, 1600],
'Fuel_Efficiency_MPG': [30, 28, 32, 27, 34, 25] })
```

Activity 2

Expected Outcome:

Built a Elastic net regression model using the dataset. Implement Elastic Net Regression using scikit-learn. Evaluate the model's performance using relevant metrics.

Steps:

1. Load and explore the dataset
2. Split the dataset into the feature and target variables
3. Create an Elastic Net Regression model
4. Evaluate the model's performance using metrics like Mean Squared Error (MSE) and R-squared
5. Compare the results with a standard linear regression model

Summary

- By adding an L2 penalty term to the loss function, Ridge Regression helps control the complexity of the model and prevents extreme coefficient values.
- Elastic Net combines both L1 (LASSO) and L2 (Ridge) regularization techniques. This method is especially useful when dealing with datasets that have multicollinearity issues.
- Case study on regularization helped to see how regularization methods can improve model performance, prevent overfitting, and enhance the generalization of models.
- Comparing the performance of models with and without regularization helps to highlight the benefits of regularization in terms of model accuracy and stability.

Session Feedback

Next Session:
Classification

THANK YOU!

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad