Data Science Learning Accelerator

Understanding Regularization in Machine Learning

Fall 2023

Sravya Vujjini

Machine learning models aim to predict or classify based on patterns in data. However, when models become overly complex, they may overfit, excelling on training data but failing to generalize to new data—hindering the primary goal of accurate real-world predictions.

The Need for Regularization

Overfitting results in poor model performance, higher errors, and reduced robustness. To combat this, regularization techniques are employed to control model complexity and mitigate overfitting.

The primary purpose of regularization is to strike a balance between fitting the training data well and ensuring the model generalizes effectively to unseen data.

There are mainly three different categories of regularization techniques:

1. Lasso Regression (L1 Regularization)
2. Ridge Regression (L2 Regularization)
3. Elastic Net Regression

Lasso Regression (L1 Regularization)

- Lasso regression, short for "Least Absolute Shrinkage and Selection Operator," is a type of linear regression that incorporates L1 regularization.

- In traditional linear regression, the goal is to find the best-fitting line by minimizing the sum of squared differences between the predicted and actual values. Lasso extends this by adding a penalty term to the linear regression's loss function, which encourages the model to have some of its coefficient values exactly equal to zero.

Pros of Lasso Regression:

1. Feature Selection: Lasso naturally performs feature selection by setting some coefficient values to zero. This can simplify the model, improve its interpretability, and remove irrelevant features, reducing the risk of overfitting.

2. Simplicity: Lasso produces a sparse model, which is beneficial when dealing with high-dimensional data where a large number of features may be present.
3. Bias-Variance Trade-Off: It helps manage the bias-variance trade-off by reducing overfitting and improving model generalization.

Cons of Lasso Regression:

1. Unstable for Highly Correlated Features: Lasso can be unstable when dealing with highly correlated features because it tends to select one feature from a group of correlated features while setting the others to zero. This can lead to unpredictability.
2. Selection of Arbitrary Features: The feature selection process might be arbitrary, meaning that different runs of Lasso may select different features when there are highly correlated variables.
3. Large Alpha May Lead to Underfitting: If the regularization strength (alpha) is too high, Lasso can underfit the data, resulting in a model that is too simplistic and performs poorly.

Ridge Regression (L2 Regularization)

- Ridge regression, also known as L2 regularization, is a linear regression technique that addresses the problem of multicollinearity and provides model stability by adding a penalty term to the linear regression's loss function.

- In linear regression, the goal is to minimize the sum of squared differences between predicted and actual values. Ridge regression extends this by adding an L2 penalty term to the loss function, which encourages the model coefficients to be smaller without necessarily becoming exactly zero.

Pros of Ridge Regression:

1. Multicollinearity Mitigation: Ridge regression effectively handles multicollinearity, a situation where predictor variables are highly correlated. It ensures the model remains stable even when dealing with correlated features.

2. Model Stability: Ridge enhances model stability, reducing the sensitivity of the model to variations in the input data, which can be crucial in cases where data is noisy or contains outliers.
3. Bias-Variance Trade-Off Management: Similar to Lasso, Ridge helps manage the bias-variance trade-off by reducing overfitting and improving generalization.

Cons of Ridge Regression:

1. No Feature Selection: Unlike Lasso, Ridge does not naturally perform feature selection. It shrinks all coefficients, but none becomes exactly zero. This means that all features are retained, which may not be desirable if some are irrelevant.
2. Limited Interpretability: The inclusion of all features can make the model less interpretable, as it retains both important and unimportant features.
3. Choice of Alpha: Selecting an appropriate value for the regularization parameter (alpha) can be a challenge. The performance of the model can be sensitive to this hyperparameter's choice.

Elastic Net Regression (Combination of L1 and L2 Regularization)

- Elastic Net regression is a hybrid approach that combines both L1 (Lasso) and L2 (Ridge) regularization techniques.
- It's designed to address the limitations and leverage the advantages of both Lasso and Ridge regression, making it a versatile method for various machine learning tasks.
- Elastic Net extends linear regression by adding a penalty term that combines both the absolute values of coefficients (L1) and their squares (L2).

Pros of Elastic Net Regression:

1. Feature Selection and Stability: Elastic Net combines the feature selection capability of Lasso with the model stability provided by Ridge. This makes it suitable for datasets with high dimensionality and multicollinearity.
2. Versatility: Elastic Net is versatile and can adapt to various data scenarios, striking a balance between feature selection and coefficient shrinkage.

3. Bias-Variance Trade-Off Management: Like Lasso and Ridge, Elastic Net helps manage the bias-variance trade-off by reducing overfitting and improving generalization.

Cons of Ridge Regression:

1. Complex Hyperparameter Tuning: Choosing appropriate values for both the regularization parameters α and λ can be complex and require thorough tuning.
2. Increased Complexity: Elastic Net can introduce additional complexity compared to Lasso or Ridge, which might be unnecessary in some cases.
3. Interpretability: While Elastic Net provides both feature selection and stability, it may still result in models that are less interpretable than simpler linear regression models.

References used :

https://chat.openai.com/chat.

OpenAI. (2023). ChatGPT (Sep 25 version) [Large language model]