Regularization

Overfitting
Overfitting occurs when a model is too complex and learns not only the underlying patterns but also the noise in the training data. This leads to excellent performance on training data but poor generalization to new, unseen data. Overfitting is characterized by low bias and high variance.

Underfitting
Underfitting happens when a model is too simple to capture the underlying structure of the data. It performs poorly on both training and test data, failing to learn the patterns in the data. Underfitting is characterized by high bias and low variance.

Regularization:
Regularization helps prevent overfitting and makes our models work better with new data. In simple terms, regularization adds a penalty to the model for being too complex, encouraging it to stay simpler and more general. This way, it's less likely to make extreme predictions based on the noise in the data. The commonly used regularization techniques are :

- Lasso Regularization - (L1 Regularization)
- Ridge Regularization - (L2 Regularization)
- Elastic Net Regularization - (L1 and L2 Regularization combined)

Lasso Regression
A regression model which uses the L1 Regularization technique is called LASSO (Least Absolute Shrinkage and Selection Operator) regression. Lasso Regression adds the "absolute value of magnitude" of the coefficient as a penalty term to the loss function(L). Lasso regression also helps us achieve feature selection by penalizing the weights to approximately equal to zero if that feature does not serve any purpose in the model.

$$\mathrm{Cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{i=1}^{m}|w_i|$$

Ridge Regression
A regression model that uses the L2 regularization technique is called Ridge regression. Ridge regression adds the "squared magnitude" of the coefficient as a penalty term to the loss function(L).

$$\mathrm{Cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{i=1}^{m}w_i^2$$

Elastic Net Regression
Elastic Net Regression is a combination of both L1 as well as L2 regularization. That implies that we add the absolute norm of the weights as well as the squared measure of the weights. With the help of an extra hyperparameter that controls the ratio of the L1 and L2 regularization.

$$\mathrm{Cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\left((1-\alpha)\sum_{i=1}^{m}|w_i| + \alpha\sum_{i=1}^{m}w_i^2\right)$$

Benefits of Regularization

- Prevents Overfitting: Regularization helps models focus on underlying patterns instead of memorizing noise in the training data.
- Improves Interpretability: L1 (Lasso) regularization simplifies models by reducing less important feature coefficients to zero.
- Enhances Performance: Prevents excessive weighting of outliers or irrelevant features, improving overall model accuracy.
- Stabilizes Models: Reduces sensitivity to minor data changes, ensuring consistency across different data subsets.
- Prevents Complexity: Keeps models from becoming too complex, which is crucial for limited or noisy data.
- Handles Multicollinearity: Reduces the magnitudes of correlated coefficients, improving model stability.
- Allows Fine-Tuning: Hyperparameters like alpha and lambda control regularization strength, balancing bias and variance.
- Promotes Consistency: Ensures reliable performance across different datasets, reducing the risk of large performance shifts.