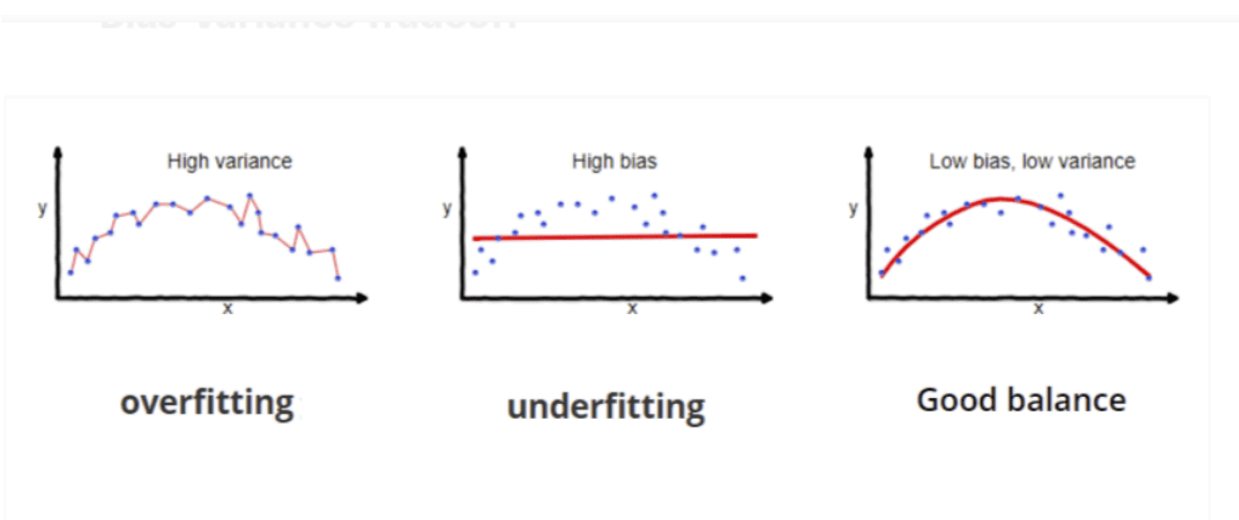
Regularization in Deep Learning: A Comprehensive Guide

Regularization is a critical concept in deep learning that helps prevent overfitting and ensures models generalize well to unseen data. It involves adding constraints or penalties to the learning process, thereby guiding the model to prefer simpler solutions that are less likely to overfit the training data. Two widely used regularization techniques are L1 and L2 regularization.

Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting. The commonly used [regularization techniques]() are :

1. Lasso Regularization – L1 Regularization
2. Ridge Regularization – L2 Regularization
3. Elastic Net Regularization – L1 and L2 Regularization



Overfitting and the Need for Regularization

Overfitting occurs when a model performs exceptionally well on the training data but poorly on test or validation data. This happens because the model learns not only the underlying patterns but also the noise in the training data. Regularization addresses this by:

1. Penalizing large model parameters.
2. Encouraging sparsity or smaller parameter values.

3. Enforcing smoother decision boundaries.

L1 Regularization (Lasso Regularization)

L1 regularization adds a penalty equal to the absolute value of the magnitudes of the weights to the loss function. The regularized loss function for L1 is expressed as:

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

Key Properties of L1 Regularization

1. **Encourages Sparsity**: L1 tends to shrink some weights to exactly zero, making it suitable for feature selection.
2. **Feature Selection**: By driving some weights to zero, L1 effectively removes irrelevant features from the model.
3. **Optimization**: Can lead to non-differentiability at zero, requiring specialized optimization techniques.

L2 Regularization (Ridge Regularization)

L2 regularization adds a penalty proportional to the square of the magnitudes of the weights to the loss function. The regularized loss function for L2 is expressed as:

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

Key Properties of L2 Regularization

1. **Penalizes Large Weights**: L2 encourages smaller weights but does not reduce any weight to exactly zero.
2. **Smoothness**: Leads to smoother decision boundaries by reducing the impact of outliers.
3. **Optimization**: Easier to optimize since the penalty term is differentiable everywhere.

Mathematical Intuition

Code: [Regularization.ipynb]