L1 and L2 Regularization in Neural Networks

Pournami P N

NIT Calicut

March 31, 2024

L1 Regularization (Lasso)

- L1 regularization adds a penalty term to the loss function that is proportional to the absolute values of the weights.
- It encourages sparsity in the weight matrix, leading to some weights being exactly zero.
- This aids in feature selection by automatically selecting a subset of the most relevant features.

Mathematical Formulation of L1 Regularization

The regularized loss function $L(\theta)$ with L1 regularization is given by:

$$L(\theta) = \text{Loss}(\theta) + \lambda \sum_{i=1}^{n} |w_i|$$

Where:

- L(θ) is the regularised objective function.
- Loss(θ) is the original loss function without regularization.
- λ is the regularization parameter.
- w_i are the individual weights in the network.

L2 Regularization (Ridge)

- L2 regularization adds a penalty term to the loss function that is proportional to the square of the weights.
- It penalizes large weight values, effectively reducing the complexity of the model and preventing overfitting.
- L2 regularization tends to produce smoother weight vectors compared to L1 regularization.

Mathematical Formulation of L2 Regularization

The regularized loss function $L(\theta)$ with L2 regularization is given by:

$$L(\theta) = \text{Loss}(\theta) + \lambda \sum_{i=1}^{n} w_i^2$$

Where:

-
- L(θ) is the regularised objective function.
- Loss(θ) is the original loss function without regularization.
- λ is the regularization parameter.
- w_i are the individual weights in the network.

Effect of L1 and L2 Regularization

- L1 regularization encourages sparsity in the weight matrix, leading to some weights being exactly zero. This aids in feature selection by automatically selecting a subset of the most relevant features.
- L2 regularization penalizes large weight values, effectively reducing the complexity of the model and preventing overfitting. It tends to produce smoother weight vectors compared to L1 regularization.
- L1 and L2 regularization are effective techniques for preventing overfitting and improving the generalization performance of neural networks by controlling the complexity of the model.
- The choice between L1 and L2 regularization depends on the desired properties of the learned model.