

Data Science 316 AF Project

Regularisation: a crucial principle in statistical learning

Andre van der Merwe
24923273

Stellenbosch University

Date of Submission: 13 May 2024

The importance of regularisation

What is regularisation?

Regularisation is a technique we use in machine learning and statistical learning to prevent overfitting and to improve a model's ability to perform accurately on unseen or new data, also known as the generalisation of a model.

Regularisation is implemented by adding a penalty term to the model's objective function (also known as the loss function or error function).

The significance of Regularisation.

The objective function of a model quantifies the error or difference between the model's predicted output and the true output.

The goal during the training of a model is to minimize this objective function. By adding a penalty term to the objective function, regularisation forces the coefficients (weights) of the model to be small. This then:

- reduces the complexity of the model
- prevents the model from overfitting

Another benefit of using regularisation is that it reduces the amount of error or noise (variance) in the model's predictions.

The impact of different strengths of the regularisation

The strength of the regularisation is controlled by the penalty term, which is a hyperparameter. Since the penalty is a hyperparameter, the person building a model, that uses regularisation, can choose the value of the penalty term.

Choosing the appropriate value for the penalty term is crucial, as it directly impacts the trade-off between model complexity and generalisation performance.

- A higher value for the penalty term results to:
 - Stronger regularisation.
 - A simpler model.
 - Models are less prone to overfitting.
- A lower value for the penalty term results to:
 - Weaker regularisation.
 - A more complex model.
 - Models may capture intricate patterns in the data.
 - Models are at higher risk of overfitting.