

STA 3180 Statistical Modelling: Model Selection

Model Selection Notes for STA 3180 Statistical Modelling

Key Concepts:

- Model selection is the process of choosing the best statistical model to represent a given set of data.
- It involves selecting the most appropriate model from a set of candidate models, based on criteria such as accuracy, parsimony, and complexity.
- The goal of model selection is to identify the model that best explains the data and makes the most accurate predictions.

Definitions:

- Model: A mathematical representation of a system or process.
- Candidate Model: A model that is being considered for selection.
- Accuracy: The degree to which a model accurately predicts the outcome of a given situation.
- Parsimony: The principle that the simplest explanation is usually the best.
- Complexity: The degree to which a model is complicated and difficult to understand.

Rules:

- When selecting a model, it is important to consider both accuracy and parsimony.
- Models should be evaluated based on their ability to accurately predict outcomes, as well as their complexity.
- It is important to choose a model that is simple enough to be understood, but complex enough to capture the nuances of the data.

Examples:

- Linear regression is a common model used in model selection. It is a simple model that can be used to predict the outcome of a given situation.
- Support vector machines are another type of model that can be used for model selection. They are more complex than linear regression, but they can provide more accurate predictions.

- Decision trees are a third type of model that can be used for model selection. They are more complex than linear regression and support vector machines, but they can provide more accurate predictions.