

# STA 3180 Statistical Modelling: Model Selection

## # Model Selection for STA 3180 Statistical Modelling

### ## Introduction

Model selection is the process of selecting a statistical model from a set of candidate models, given data. It is an important part of statistical modelling, as it helps to identify the best model that fits the data and provides the most accurate predictions. Model selection involves choosing the most appropriate model from a set of candidate models, based on the data available. This process can be done manually by examining the data and making a judgement, or it can be done using automated techniques such as cross-validation or AIC/BIC.

### ## Key Concepts

\* \*\*Candidate Models\*\*\*: A set of models that are considered for selection.

\* \*\*Data\*\*\*: The information used to make the selection.

\* \*\*Cross-Validation\*\*\*: A method of evaluating models by dividing the data into training and test sets, and then using the test set to evaluate the performance of the model.

\* \*\*AIC/BIC\*\*\*: Akaike Information Criterion and Bayesian Information Criterion, two methods of evaluating models based on their complexity and accuracy.

### ## Definitions

\* \*\*Model Selection\*\*\*: The process of selecting a statistical model from a set of candidate models, given data.

\* \*\*Candidate Models\*\*\*: A set of models that are considered for selection.

\* \*\*Data\*\*\*: The information used to make the selection.

\* \*\*Cross-Validation\*\*\*: A method of evaluating models by dividing the data into training and test sets, and then using the test set to evaluate the performance of the model.

\* \*\*AIC/BIC\*\*\*: Akaike Information Criterion and Bayesian Information Criterion, two methods of evaluating models based on their complexity and accuracy.

### ## Coding Examples

#### ### Example 1: Cross-Validation

Start of Code

```
// Split the data into training and test sets
let trainData = data.slice(0, data.length * 0.8);
let testData = data.slice(data.length * 0.8);
// Train the model on the training data
let model = train(trainData);
```

```
// Evaluate the model on the test data
let score = evaluate(model, testData);
End of Code
```

### ### Example 2: AIC/BIC

Start of Code

```
// Calculate the AIC/BIC for each model
let aicScores = [];
let bicScores = [];
for (let model of models) {
    let aicScore = calculateAIC(model);
    let bicScore = calculateBIC(model);
    aicScores.push(aicScore);
    bicScores.push(bicScore);
}
// Select the model with the lowest AIC/BIC score
let bestModelIndex = aicScores.indexOf(Math.min(...aicScores));
let bestModel = models[bestModelIndex];
End of Code
```

### ## Practice Multiple Choice Questions

Q1. What is model selection?

A. The process of selecting a statistical model from a set of candidate models, given data.

Q2. What is cross-validation?

A. A method of evaluating models by dividing the data into training and test sets, and then using the test set to evaluate the performance of the model.