AIC and BIC: Mathematical Formulation and Intuition

Jerome Chou

February 24, 2025

Introduction

- Model selection is crucial in statistical learning and machine learning.
- The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are widely used to compare models.
- These criteria balance model fit and complexity.

Akaike Information Criterion (AIC)

Definition:

$$AIC = -2 \ln L(\hat{\theta}|\mathcal{D}) + 2k \tag{1}$$

where:

- $L(\hat{\theta}|\mathcal{D})$ is the maximum likelihood of the model given data $\mathcal{D}$,
- k is the number of estimated parameters.

Intuition:

- The first term rewards model fit (higher likelihood is better).
- The second term penalizes model complexity (to prevent overfitting).

Bayesian Information Criterion (BIC)

Definition:

$$BIC = -2 \ln L(\hat{\theta}|\mathcal{D}) + k \ln n \tag{2}$$

where:

- n is the sample size.

Intuition:

- Similar to AIC but with a stronger penalty for complexity.
- As n grows, the penalty term increases, favoring simpler models.

Comparison: AIC vs. BIC

- AIC aims to minimize prediction error (better for out-of-sample accuracy).
- BIC is based on Bayesian probability and prefers the true model as $n \to \infty$.
- BIC penalizes complexity more heavily than AIC.
- Rule of thumb:
  - Use AIC when the goal is prediction.
  - Use BIC when the goal is model selection with a true underlying model.

Example Calculation

Given:

- Sample size: $n = 100$
- Log-likelihood: $\ln L = -250$
- Number of parameters: $k = 5$

Compute:

$$AIC = -2(-250) + 2(5) = 500 + 10 = 510$$
$$BIC = -2(-250) + 5 \ln 100 = 500 + 5(4.605) = 523.03$$

Interpretation:

- Lower values indicate a better model.
- If comparing multiple models, choose the one with the smallest AIC/BIC.

Summary

- AIC and BIC are tools for model selection.
- AIC is focused on minimizing prediction error.
- BIC is more conservative and prefers simpler models when n is large.
- Neither is perfect; always consider domain knowledge.