Model Selection Criteria: AIC, AICc, and BIC

Tanmay Chopra

Here are clear, concise explanations of **AIC**, **AICc**, and **BIC**, three common criteria used for **model selection** in statistics and machine learning:

AIC — Akaike Information Criterion

Purpose: Choose the model that best explains the data with as few parameters as possible.

Formula:

$$\text{AIC} = -2\ln(L) + 2k$$

Where:

- L = maximum likelihood of the model
- k = number of estimated parameters

Interpretation:

- Lower AIC = better model (relative comparison)
- Balances goodness of fit and model complexity
- Good for predictive performance

Limitation:

- Can favor more complex models when sample size is small

AICc — Corrected Akaike Information Criterion

Purpose: AIC adjusted for **small sample sizes**.

Formula:

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}$$

Where:

- n = sample size

When to use:

- When $n/k < 40$ (sample size is not large relative to number of parameters)

Interpretation:

- Always $\geq$ AIC
- Prevents selecting overly complex models with small samples

---

BIC — Bayesian Information Criterion

Purpose: Model selection based on approximate Bayesian reasoning.

Formula:

$$\text{BIC} = -2\ln(L) + k\ln(n)$$

Differences from AIC:

- Penalizes complexity more strongly because of the $(\ln(n))$ term
- Tends to choose simpler models when sample size is large

Interpretation:

- Lower BIC = better model
- More conservative than AIC

Quick Comparison Table

Criterion	Penalizes Complexity	Best For	Behavior
AIC	Moderate $(2k)$	Prediction accuracy	Often selects more complex models
AICc	Higher at small n	Small-sample problems	Safest when n is limited
BIC	Strong $(k \ln n)$	Bayesian model selection, large n	Favors simpler models