| CS57300 Homework | — Homework 4 , Problem 1 — | (April 8, 2016) |
|---|---|---|
| Shaobo Fang ( 0024260324 ) | | |

**Q1:**
Describe the difference between BIC and AIC for model selection. When should you choose BIC over AIC? And vice-versa?
**A:**

The AIC (Akaike Information Criterion) is defined as:

$$AIC = \frac{2}{N} \ln P[t = f(x_0; \hat{\theta})] + 2\frac{d}{N}$$

The BIC (Bayesian Information Criterion) is defined as:

$$BIC = \frac{2}{N} \ln P[t = f(x_0; \hat{\theta})] + \frac{d}{N} \log N$$

where the likelihood function is:

$$P[t = f(x_0; \hat{\theta})]$$

**Difference:**
AIC: AIC does not assume model is correct. This is better for small number of samples N. And it is not guarantee to converge to "correct model" (as $N =\rightarrow \infty$ model bias $\geq 0$)
BIC: BIC assumes model is correct. This converges to correct parameterization ($\theta^*$) of model (as $N =\rightarrow \infty$ model bias $= 0$).
The case where BIC is better than AIC are those that we have the correct model but need to find the best parameters. As the assumption of BIC is the model is correct and it will converge to the correct parameterization.
Vice versa, in real world scenarios, very likely we will not have the correct model (and probably there does not exist a definitive probability model to govern the observation). Hence we should choose AIC if we are not sure if the prior model is correct. AIC is also better than BIC if with small number of observations. In another sense, AIC is better for real world scenarios where model is unknown and not enough observations.

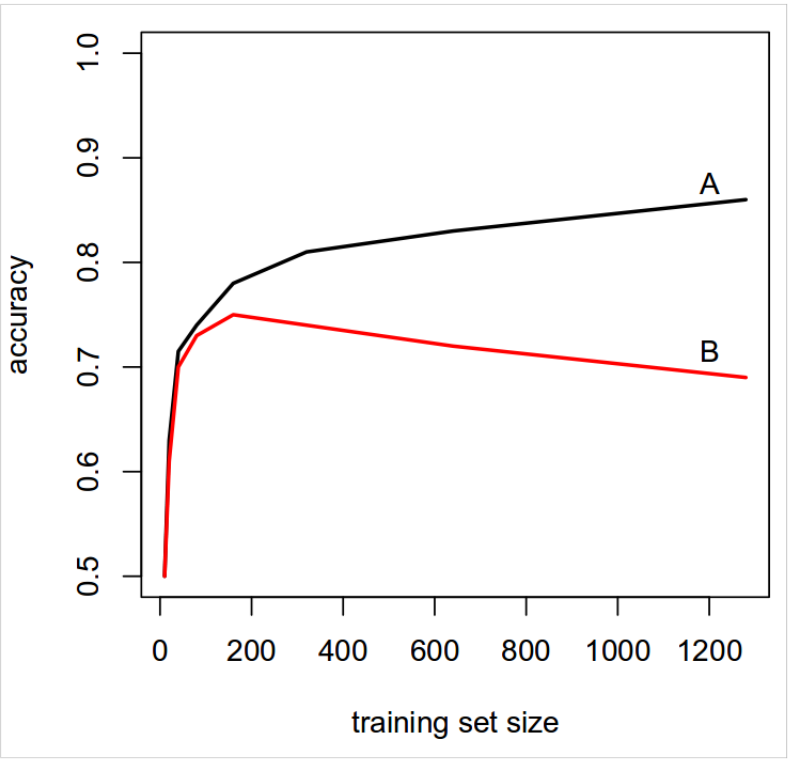


Figure 1: The Training - Testing Curve

| | | |
|---|---|---|
| CS57300 Homework | — Homework 4 , Problem 2 — | (April 8, 2016) |
| Shaobo Fang ( 0024260324 ) | | |

**Q2:**
(i) Which one is the training dataset and which one is the testing dataset.
(ii)Is the model M over-fitting? Why or why not? Explain how the data in the figure supports your conclusion.
**A:**

(i). The training set is A, as the accuracy of the set always increase with the size of the training set increasing.
The testing set is B, as the accuracy of the set will decrease with the size increasing after one point.

(ii). The model is over-fitting. Because with more training data although the accuracy on training dataset increased the accuracy decreased on testing dataset. Clearly indicate on the figure that the accuracy of training set and testing does not increase simultaneously (this behaviour starting around the size of 200).