Bayesian model selection

Dr. Merlise Clyde

Duke
UNIVERSITY

model selection criteria

Bayesian information criterion (BIC)

- $-2 \times \log(likelihood) + \log(n) \times \#parameters$
- smaller is better
- $n \log(1 - R^2) + \log(n) \times \#\text{parameters}$
- trade-off goodness of fit with model complexity

backwards elimination with BIC

backwards elimination with BIC

step	model	dropped	BIC
FULL	kid_score ~ hs + iq + work + age		2541.1

backwards elimination with BIC

step	model	dropped	BIC
FULL	kid_score ~ hs + iq + work + age		2541.1
STEP 1	kid_score ~ hs + iq + work	[-age]	**2535.4**
	kid_score ~ hs + iq + age	[-work]	2536.2
	kid_score ~ iq + work + age	[-hs]	2539.9
	kid_score ~ hs + work + age	[-iq]	2614.1

backwards elimination with BIC

step	model	dropped	BIC
FULL	kid_score ~ hs + iq + work + age		2541.1
STEP 1	kid_score ~ hs + iq + work	[-age]	**2535.4**
	kid_score ~ hs + iq + age	[-work]	2536.2
	kid_score ~ iq + work + age	[-hs]	2539.9
	kid_score ~ hs + work + age	[-iq]	2614.1
STEP 2	kid_score ~ hs + iq + work		2535.4
	kid_score ~ hs + iq	[-work]	**2530.6**
	kid_score ~ iq + work	[-hs]	2535.0
	kid_score ~ hs + work	[-iq]	2608.8

backwards selection continued

step	model	dropped	BIC
FULL	kid_score ~ hs + iq + work + age		2541.1
STEP 1	kid_score ~ hs + iq + work	[-age]	2535.4
STEP 2	kid_score ~ hs + iq	[-work]	**2530.6**
STEP 3	kid_score ~ hs + iq		2530.6
	kid_score ~ hs	[-iq]	2531.7
	kid_score ~ iq	[-hs]	2604.0

best model kid_score ~ hs + iq

best adjusted-R^2 kid_score ~ hs + iq + work

estimates under reference prior

reference prior

- uniform distribution for coefficients
- $p(\sigma^2) \propto 1/\sigma^2$

reference posterior distribution

	posterior mean	sd	2.5%	97.5%
(intercept)	19.78	6.29	7.43	32.14
hs	5.95	2.21	1.60	10.30
iq	0.56	0.06	0.44	0.68

other criteria

- $-2 \times \log(likelihood) + k \times \#parameters$
- pick model with highest posterior probability
- pick best predictive model
- include costs associated with using models

next video

- model uncertainty