HW6

Nical Karim

1. Say whether the following statements are true or false and explain why.

   (a) For *any* set of predictor variables, the larger the number of predictor variables in the model, the larger the R^2.

Generally true, it could also stay the same. But this doesn't mean the model is better as it could be overfit. Adjusted R^2 can account for this.

   (b) For model of the same size (fixed p), their C_p, AIC_p, BIC_p values are monotonically increasing in terms of SSE_p.

True SSE increases the fit of the model worsens, and so C_p, AIC, BIC will penalize for these higher errors so yes they will increase

   (c) Compared with AIC, BIC criterion tends to select smaller models because it puts higher penalties on model size.

True, BIC has log(n) term while AIC uses constant term of just 2

   (d) The best subsets procedure is guaranteed to find the "best" model under a given criterion.

True, it explores every possible combination of predictors to optimize based on the given criterion so it must find the best single combination

2. Practice model selection on an example data set. Data set "HW6Q2.txt" contains 4 variables with the response variable Y on the first column followed by 3 predictor variables. We consider all first-order models.

   (a) How many first-order models are there?

   3 models with 1 predictor

   3 models with 2 predictors

   1 model with 3 predictor.

   7 models

   (b) Among all the first-order models, report the "best" models according to each of the following criteria: $R^2_{adj,p}$, AIC_p, BIC_p, C_p, $PRESS_p$, as well as their corresponding values according to the criterion.

R^2_{adj} : model $[X_1, X_2, X_3]$: 0.553

AIC_p : Model $[X_1, X_3]$: 159.8

BIC_p : Model $[X_1, X_3]$: 165.5

Mallow C_p : Model : $[X_2, X_3]$: 0.578

PRESS : Model : $[X_1, X_2]$: 71.753

(c) Using AIC_p, select the best overall model of any size. Using this model, check for influential points using Cook's Distance. If there are any, print them out.

Best AIC model : $[X_1, X_3]$

Influential points :

	Y	X_1	X_3
2	10.5	10.96	-1.5
46	14.5	-3.77	14.30
48	16.75	-3.65	16.77

3. For the data set IceCreamConsumption.csv and consider y=cons with predictors income, price, and temp.

(a) List all the possible models from this data set (without interactions or higher powers).

See notebook for full code.

Model 1: Predictors = ['income']
Model 2: Predictors = ['price']
Model 3: Predictors = ['temp']
Model 4: Predictors = ['income', 'price']
Model 5: Predictors = ['income', 'temp']
Model 6: Predictors = ['price', 'temp']
Model 7: Predictors = ['income', 'price', 'temp']

(b) Calculate the adjusted R^2 and C_p for all the models, make a summary table with four columns: Number of predictors, R_a^2 values, C_p values, Predictors in the model.

Summary:

	Number of Predictors	Adjusted R^2	Mallow's C_p	Predictors
0	1	-0.033334	0.933364	[income]
1	1	0.034082	-0.954282	[price]
2	1	0.587365	-16.446210	[temp]
3	2	-0.001257	1.033942	[income, price]
4	2	0.679989	-17.359708	[income, temp]
5	2	0.605619	-15.351717	[price, temp]
6	3	0.686570	-15.850822	[income, price, temp]

(c) Based on the table above, which model is selected by R_a^2? By C_p?

R^2 adj : model: [income, price, temp] : 0.687

Full model

Mallows Cp: model [price] : 0.933

Cp closest to 1

(d) For the two models in part c., calculate the AIC and BIC values. Based on AIC and BIC, what's your final choice of model?

Summary Table

	Model	Number of Predictors	AIC	BIC
0	[income, price, temp]	3	-109.238872	-103.634082
1	[income]	1	-75.226523	-72.424129

AIC and BIC more negative in full model suggests it is a better fit.

Choose model

[income, price, temp]

(e) Is there a difference in the size of the model selected by AIC and BIC? If yes, state which is more parsimonious and explain why this difference exists.

Here, there is not but in general BIC does prefer simpler smaller models since it punishes more for complexity compared to AIC.