Additional note 4

Lecture 11

1. **Revision of forward and backward selection in L10**

   - What are the R commands for backward and forward selection?

     Answer: In a backward selection, we use `drop1` to drop one variable at a time from the current model and then update the current model using `update`. For a forward selection, we use `add1` to add one variable at a time to the current model and then update the current model using `update`.

   - What are the selection criteria for adding or dropping variables?

     Answer: In the case of using F-test as the selection criterion, we drop the variable x if $p_x > p_{\text{out}}$ for backward selection and we add the variable x if $p_x < p_{\text{in}}$ for forward selection. The cutoffs p_{out} or p_{in} need to be set or adopt the defaults. We will see other selection criteria, AIC and BIC.

   - Why should we add or drop just only one variable at a time?

     Answer: It is because some x variables can be highly correlated (multicollinearity) and so adding or dropping one variable can greatly affect the p-values of the remaining variables in the model. In other words, the selected model is very local and conditional and different cutoffs, selection criteria and selection methods may end up with different selected models.

2. **Stepwise selection**

   - Stepwise procedures allow adding of one variable among those outside the current model and dropping of one variable among those inside the current model at different steps of the selection process. The aim is to go through the set of models as much as possible to select the best model. Now we need both p_{out} and p_{in} for dropping and adding variables.

   - Note that $p_{\text{out}} > p_{\text{in}}$. Let the p-value of a x variable be $p_x = 0.15$.

     If $p_{\text{out}} = 0.2$ and $p_{\text{in}} = 0.1$:
     If x is outside the model and is in the top of the list, x is not added since $0.15 > 0.1$.
     If x is inside the model and is in the top of the list, x is not dropped since $0.15 < 0.2$.

     If swapped, ie $p_{\text{out}} = 0.1$ and $p_{\text{in}} = 0.2$:
     If x is outside the model and is in the top of the list, x is added since $0.15 < 0.2$.
     Once x is inside the model and is in the top of the list, x is dropped since $0.15 > 0.1$. So the add and drop are repeated in a loop.

   - Other selection criteria apart from F-test are R_a^2, AIC, C_p and BIC. The general format is

     $$\begin{aligned} \text{Selection criterion} &= \text{Model fit measure} + \text{Model complexity measure} \\ &\overset{AIC,BIC}{=} \text{-2 loglikelihood} + kp \end{aligned}$$

     where the second term penalise complex model, k is the penalty weight ($k = 2$ for AIC and $k = log(n)$ for BIC) and p is the number of model parameters.

   - AIC and BIC differ from other criteria which use mainly RSS (a nonparametric measure comparing only observed with predicted) to measure model fit whereas AIC and BIC use loglikelihood which is the sum of log density and so is a parametric measure. As there is no p-value for HT associated with AIC and BIC, they can be used in non-nested model for model comparison applied to the same data set. As the sample size n increases, the magnitude of log-likelihood also increases and so the model complexity penalty $2p$ in AIC becomes relatively light. Hence BIC changes $2p$ to $\log(n)p$ to allow for the sample size n.

- In each step, all variables outside the current model have + to be added and those inside the current model have − to be dropped. Then they together with the current model <none> are ranked in descending order of AIC or BIC (since higher indicates worse fit). As the steps move on, the current model <none> improves and so is ranked higher and higher until it is on the top which means no further improvement of AIC or BIC for adding or dropping of variables and the process will terminate.

Lecture 12

1. **Revision questions of L10,L11**

   - What other criteria apart from F-test can be considered in model selection?

     Answer: R_a^2, C_p, AIC and BIC.

   - Should AIC and BIC be applied to nested model only? Are there p-values to set?

     Answer: No need to set p-values. They are not used for hypothesis testing of nested models so they can be used to compare any group of models from the same data set.

   - How to ensure parsimonious model? What are the penalty weight for AIC and BIC?

     Answer: To ensure parsimonious model, one should penalise the model fit with model complexity which relates to number of parameter p and possibly sample size n. We can see all F-test statistic, R_a^2 and C_p include p and n but their model fits are based on RSS. For AIC and BIC, the model complexity is a separate measure taking the form kp in which $k = 2$ for AIC and $k = \log(n)$ for BIC.

2. Read additional slides on P.17-19 for a 3-dimensional plot of response surface and some comments.

---

last adjustments: March 2, 2021 by JC