Spring 2022

Quantitative Analysis
Chung-Ming Kuan
Jui-Chung Yang

Final

This is an 1-hour open book exam. Please make sure that your answers are legible and clearly labeled. All the results from Prof. Kuan's and Prof. Yang's lecture slides and HW could be used directly in your answers, while other results must be proved and explained clearly. Good luck!

1. (15 points)

    Answer the following questions with "True" or "False" and briefly explain it. Answers without justifications will receive *no* points. All notations are defined the same as in our lecture slides.

    (a) The LASSO regression will predict 0 when the tuning parameter, λ, goes to infinity.

    (b) Suppose that we fit our data $(x_i, y_i)_{i=1}^{n}$ using local regression with tricubic weights and span $s = 50\%$. This requires fitting local regression model *twice* to obtain the fitted curve for our data.

    (c) Suppose we fit data with forward stepwise selection, backward stepwise selection, and best subset selection procedures, and obtain three respective selected models. Two of them should be identical.
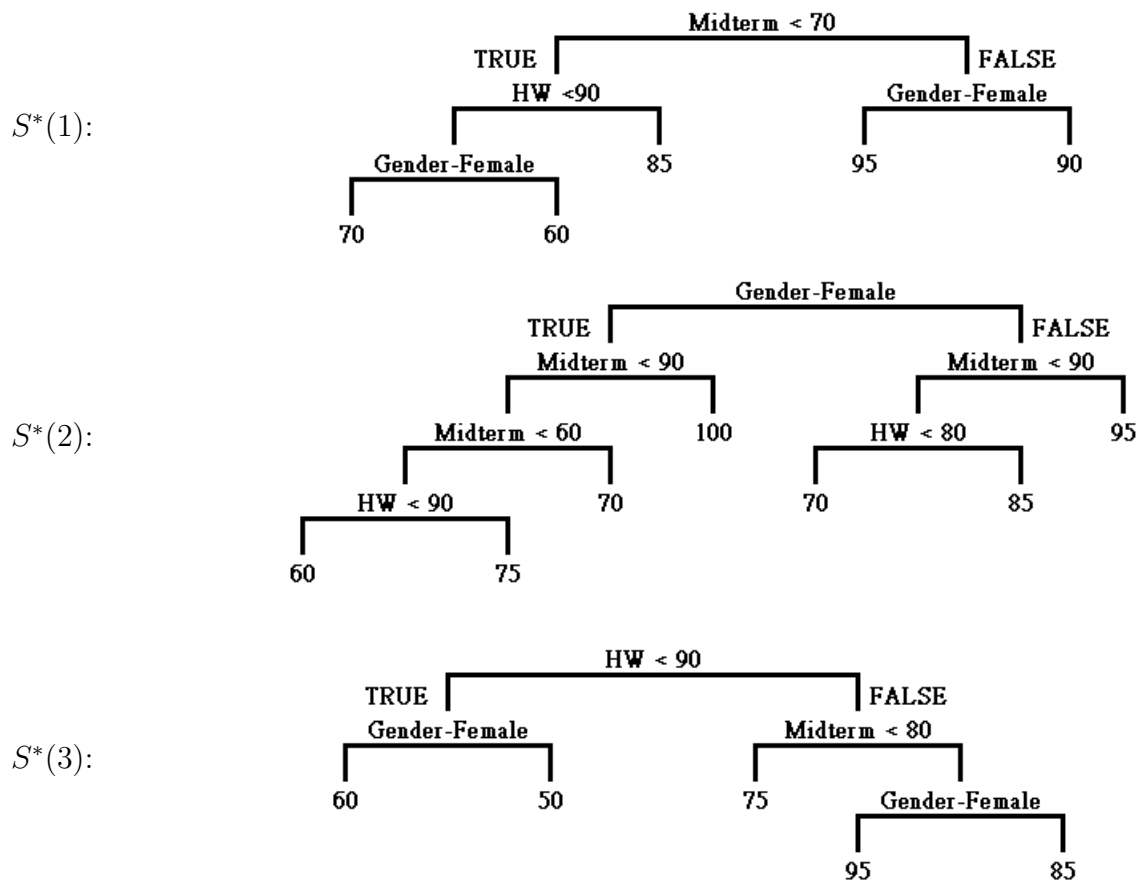
2. (5 points)

    We calculate testing MSEs for one particular model and data set with three different cross validations: LOOCV, 10-fold CV, and the validation set approach, each for 10 times. The resulting tesing MSEs are shown below:

Method A	5.8021	5.8021	5.8021	5.8021	5.8021	5.8021	5.8021	5.8021	5.8021	5.8021
Method B	5.8523	5.8266	5.8295	5.7586	5.7909	5.8205	5.7719	5.8005	5.8231	5.7831
Method C	5.7004	6.5871	6.0887	6.3250	5.3031	5.5726	6.0595	5.8611	5.7672	4.8724

    Which cross validation do method A, B and C respectively correspond to? Briefly explain your answer.

Test MSE computation

Decoding the trees and computing OOB predictions for each observation:

obs.1 (Midterm=80, HW=85, Female, Final=70), OOB: $S^*(3)$
- $S^*(3)$: HW<90 TRUE → Gender-Female TRUE → **60**
- Prediction = 60, error^2 = $(70-60)^2 = 100$

obs.2 (Midterm=75, HW=85, Female, Final=75), OOB: $S^*(1)$
- $S^*(1)$: Midterm<70 FALSE → Gender-Female TRUE → **95**
- Prediction = 95, error^2 = $(75-95)^2 = 400$

obs.3 (Midterm=95, HW=95, Female, Final=90), OOB: $\{S^*(1), S^*(3)\}$
- $S^*(1)$: Midterm<70 FALSE → Gender-Female TRUE → 95
- $S^*(3)$: HW<90 FALSE → Midterm<80 FALSE → Gender-Female TRUE → 95
- Prediction = $(95+95)/2 = 95$, error^2 = $(90-95)^2 = 25$

obs.4 (Midterm=65, HW=65, Male, Final=60), OOB: $\{S^*(1), S^*(2)\}$
- $S^*(1)$: Midterm<70 TRUE → HW<90 TRUE → Male → 60
- $S^*(2)$: Male → Midterm<90 TRUE → HW<80 TRUE → 70
- Prediction = $(60+70)/2 = 65$, error^2 = $(60-65)^2 = 25$

obs.5 (Midterm=55, HW=75, Male, Final=60), OOB: $S^*(2)$
- $S^*(2)$: Male → Midterm<90 TRUE → HW<80 TRUE → **70**
- Prediction = 70, error^2 = $(60-70)^2 = 100$

$$\text{Test MSE} = \frac{100 + 400 + 25 + 25 + 100}{5} = \frac{650}{5} = 130$$