Assignment #5— Due Monday, 6 Dec.

*This homework covers Chapter 9. Submit your homework on Canvas or send it to our TA, Mr. LYU Zhongyuan (zlyuab@connect.ust.hk).

*No late homework will be accepted for credit.

*Append the R codes you used to your submission. *If the problem does not need R or is not explicitly stated to complete in R, then you should just do it by hand with a calculator.*

*In case of rounding error, keep 3 figures after the decimal point.

Problem 1 (Use R) The purpose is to predicting whether a patient has breast cancer based on 9 predictors (features of the tumor) as follows:
Cl.thickness: Clump Thickness
Cell.size: Uniformity of Cell Size
Cell.shape: Uniformity of Cell Shape
Marg.adhesion: Marginal Adhesion
Epith.c.size: Single Epithelial Cell Size
Bare.nuclei: Bare Nuclei
Bl.cromatin: Bland Chromatin
Normal.nucleoli: Normal Nucleoli
Mitoses: Mitoses
The response is the variable *class* taking two possible values *benign* and *malignant*. The data was collected by Dr. Wolberg from his clinical cases, containing 683 observations, stored in the file *BreastCancer.txt*.

(a) Fit a logistic regression model using all the predictor variables, report the summary and interpret your results. Is the model significant?

(b) Based on part (a), construct the 95% confidence interval for $\beta_{Cl.thickness}$ and test the hypothesis

$$H_0 : \beta_{Cell.shape} = 0 \quad \text{versus} \quad H_1 : \beta_{Cell.shape} \neq 0$$

at the significance level 0.10.

(c) Use the predictors *Cl.thickness, Cell.shape, Marg.adhesion, Bare.nuclei, Bl.cromatin* to fit a logistic regression model, report the summary and compare it to the full model in part (a).

(d) Given part (c) and a new patient with *Cl.thickness=6, Cell.shape=3, Marg.adhesion=8, Bare.nuclei=2, Bl.cromatin=5*, what is the probability of this patient's tumor being *benign*?

(e) Compare the model in part (c) and all its subset models (i.e., models only using variables from *Cl.thickness, Cell.shape, Marg.adhesion, Bare.nuclei, Bl.cromatin*), which of them has the smallest AIC? Report your conclusions.