CSC591: Foundations of Data Science

HW4: Feature Selection, Resampling, Bootstrap

Released:
Due:

Student Name:
Student ID:

Notes

- Filename: Lastname_StudentID.pdf (only pdf).
- You can also submit scanned hand-written solution (should be legible, TA's interpretation is final).
- This h/w is worth **4**% of total grade.
- *Answer all questions. A subset of the questions will be graded by the TAs* (these questions will be decided by the instructor).
- All submission must be through Moodle (you can email to TA with cc to Instructor – only if these is a problem – if not received on time, then standard late submission rules apply)
- No makeups or bonus; for regarding policies, refer to syllabus and 1st day lecture slides.
- You are encouraged to do research, study online materials; discuss with fellow students; BUT ANSWERS SHOULD BE YOUR OWN. Any kind of copying will result in 0 grade (minimum penalty), serious cases will be referred to appropriate authority.
- All questions of this h/w require hand calculations using the formulas that you learned from the course materials. You can use any regular calculator. [Note: It is important to do it by hand as you are expected to do it in exam; so, this gives you practice; show all calculations in tabular form]

Q1. Information theory (20 points)

Using the data provided in the following table, **rank the attributes** based on information gain (lowest to highest). Also note that the first column is serial number and last column is classification attribute
You can assume 0 * log2(0) = 0

Example#	Type of call	Language frequency	Ticket type	Age	Decision Attribute
E1	Local	Fluent	Long	Very young	Buy
E2	Long Dist.	Fluent	Local	old	Buy
E3	Long Dist.	Not Fluent	Short	very old	Buy
E4	Intern.	Accent	Long	very old	Buy
E5	Local.	Fluent	Short	middle	Buy
E6	Local	Not Fluent	Short	very young	Not Buy
E7	Intern.	Accent	Short	middle	Not Buy
E8	Intern.	Foreign	Long	young	Not Buy
E9	Local	Not Fluent	Long	middle	Not Buy

Q2. Dimensionality Reduction(20 points)
Find principal components for the following 2-dimensional data:

X_1	X_2
10	-3
9	-1
8	-2
11	-4
7	0

Q3. Linear Model Selection (10 points)
(a) Of best subset and stepwise (forward and backward) selection methods on a given dataset consisting of “p” covariates, answer the following:

Which of the three models with “k” predictors (where k << p) has the smallest training RSS and explain why?

(b) What are the similarities and dissimilarities between ridge regression and LASSO? Which method do you prefer and why?