Please provide clear and detailed answers for the midterm problems. Any unclear proof or interpret will not be accepted. No discussion with anyone except your instructor. Deadline: (1/11 (Friday) 14:00 -15:00 at A432)

(1) (20 pts) Please explain what the ridge and Lasso regression are, and the difference between the least square estimator, the ridge estimator and Lasso estimator.

(2) (Cross-validation) (20 pts) For $n$ observed data $(X_{n\times p}, Y_{n\times 1})$, assume theta $\hat{Y}_{i(i)}$ is the esitmated expected value based on data except $i$ the case. Please show $Y_i - \hat{Y}_{i(i)} = \frac{e_i}{1-h_{ii}}$, where $e_i$: the residual of $i$th data based on all data information. $h_i$ : the $i$ diagonal element in the Hat matrix.

(3) A manager would like to increase the return on an investment, which depends on 17 financial products. But, some of financial products can not affect the return. In the attached data set including 2000 data, the first 17 columns show the investment for each product, and the last one is the corresponding return.

(a) (25 pts) Please analyze the data (It is required to provide ANOVA table.)?

(b) Assume that the budget (US 100000) is limited. Based on your analysis results in (a) and (b),

1. (15 pts) please make your comparison on two different strategies (20000,30000,0,40000,10000,0....,0) and (50000,0,20000,0,30000,0....,0).
2. (20 pts) Also, please discuss the prediction of their return throught the theoretical way.