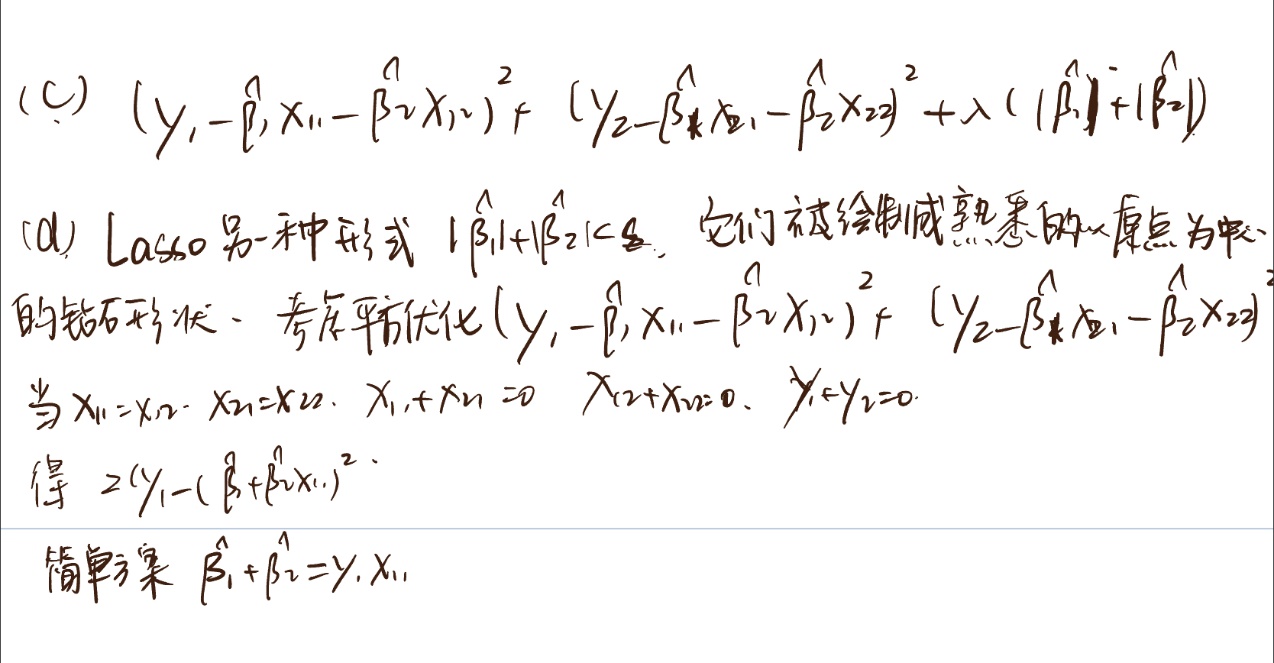
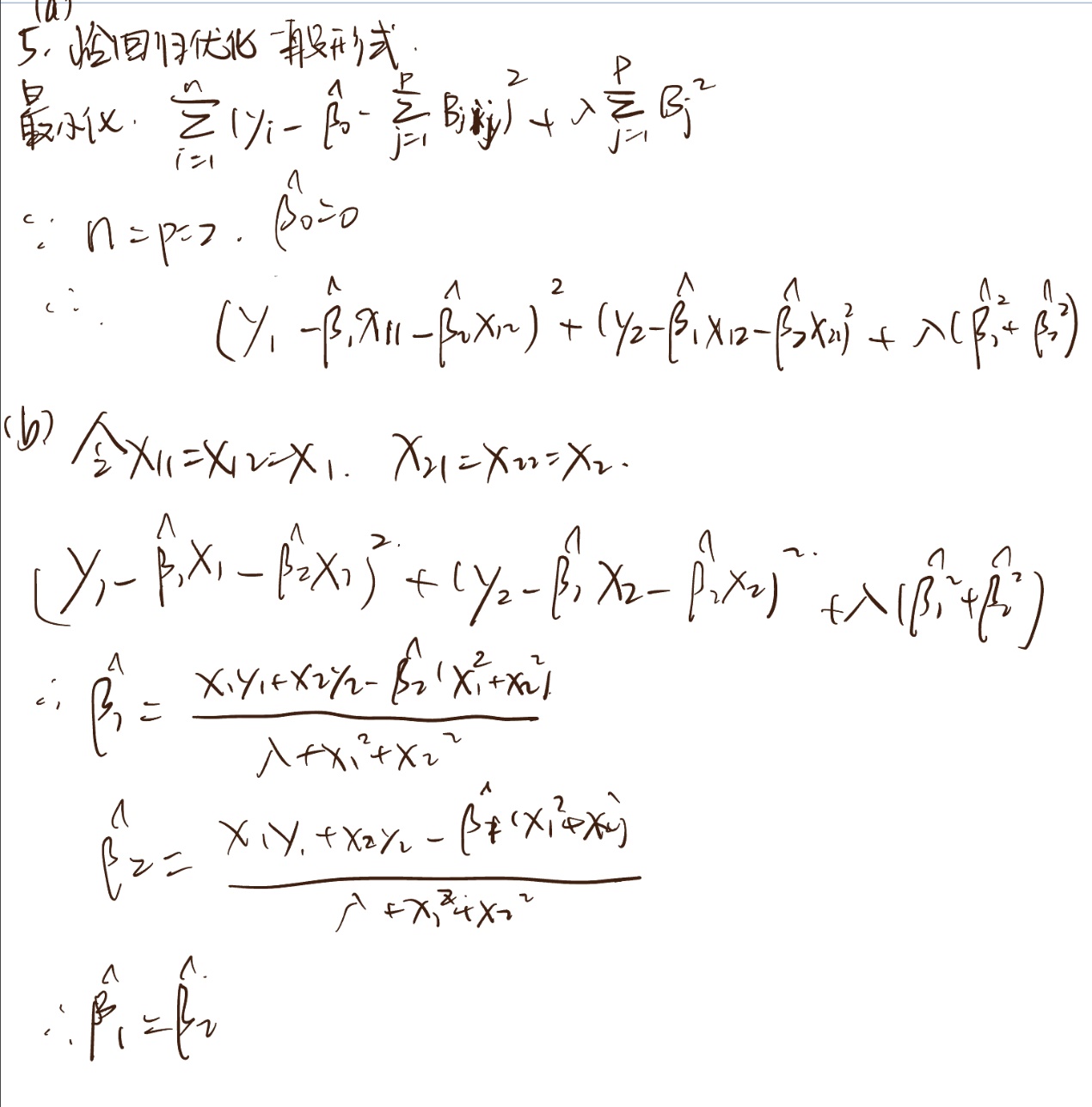
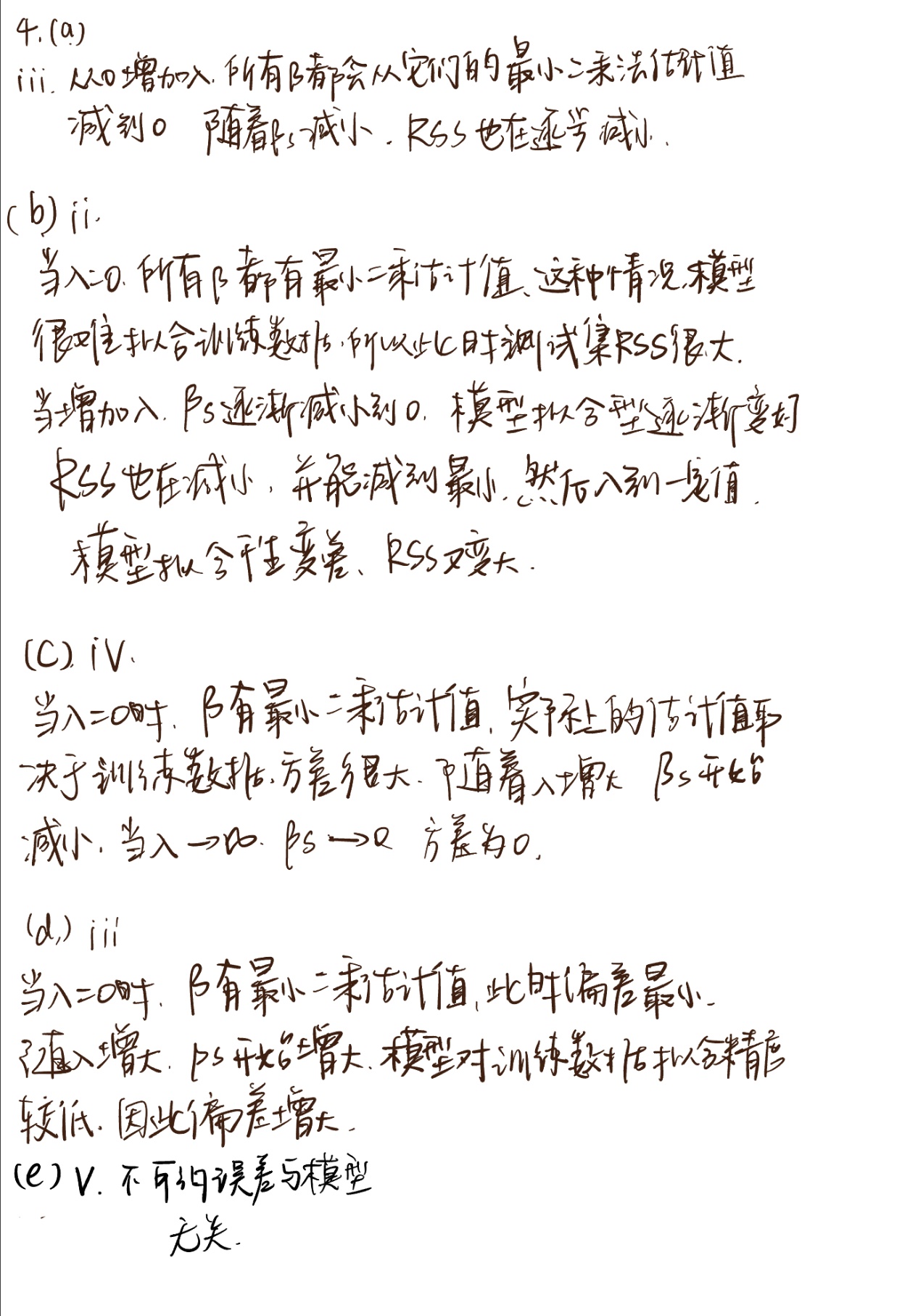
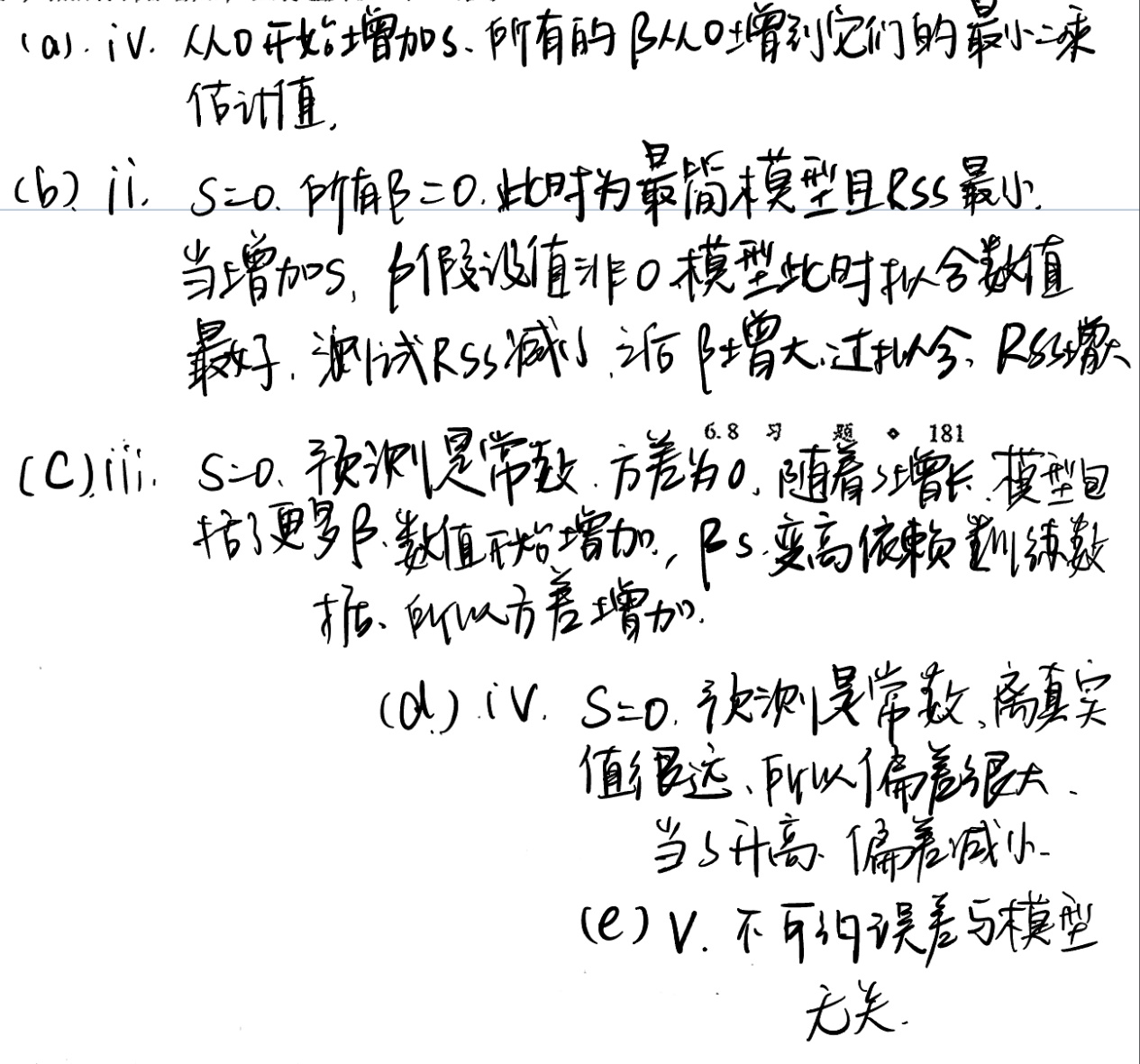
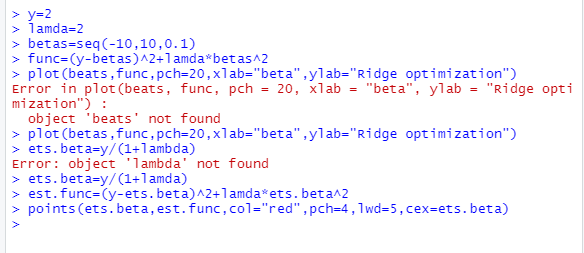
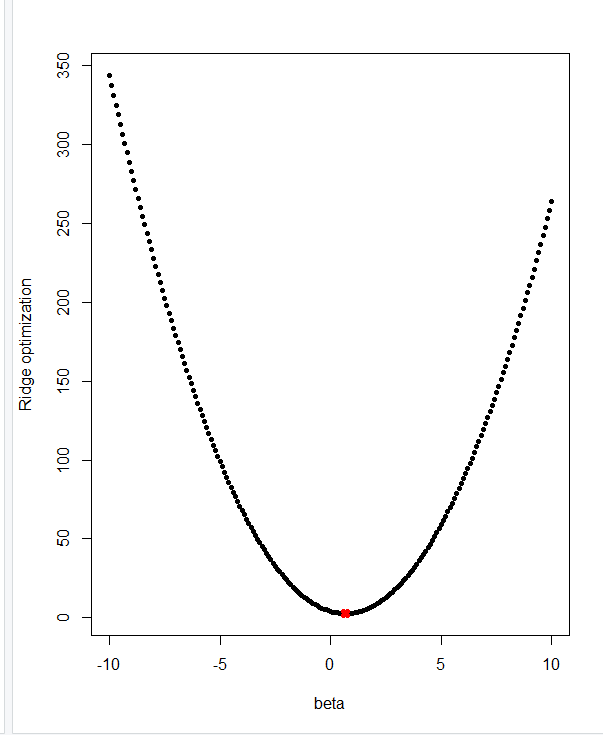
3.

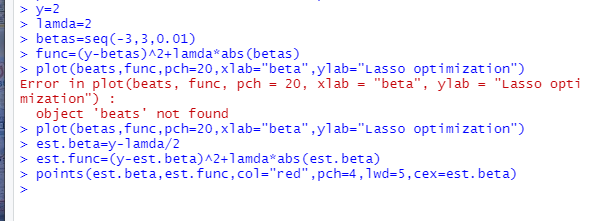


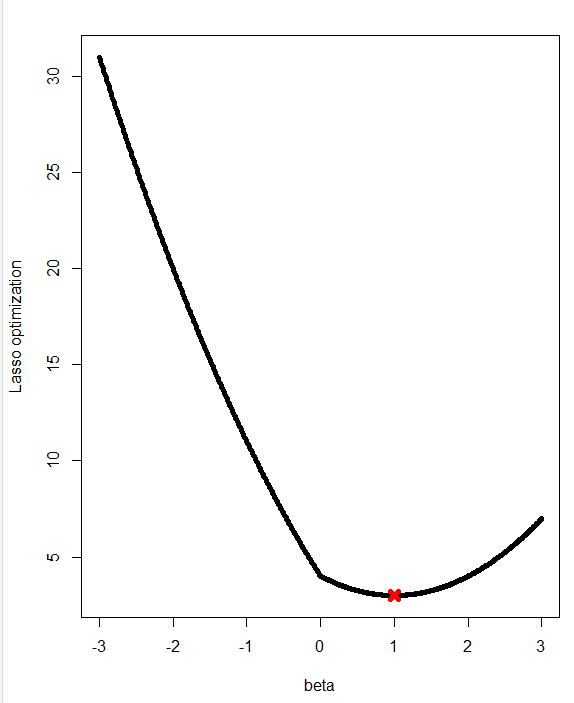
6（a）





6（b）





Raw=T 是正常的多项式（不是正交的）

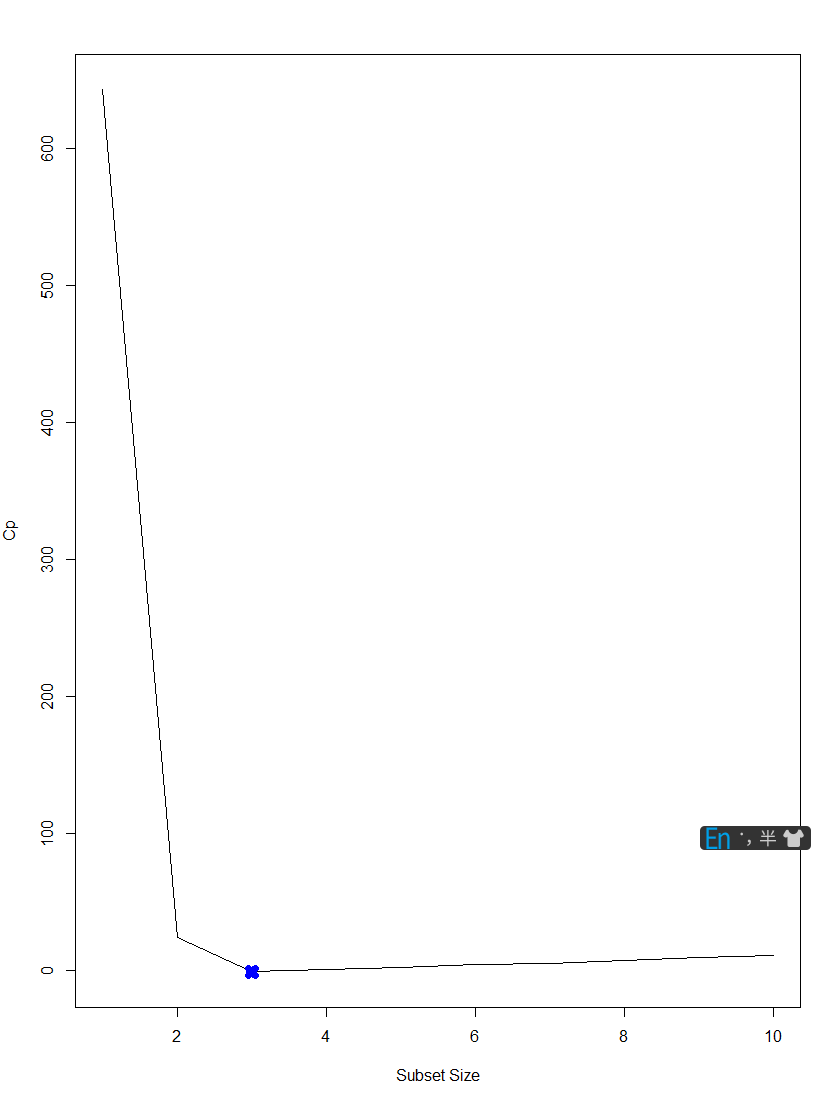
Regsubsets是最优子集选择法，nvmax是最多选择的特征个数

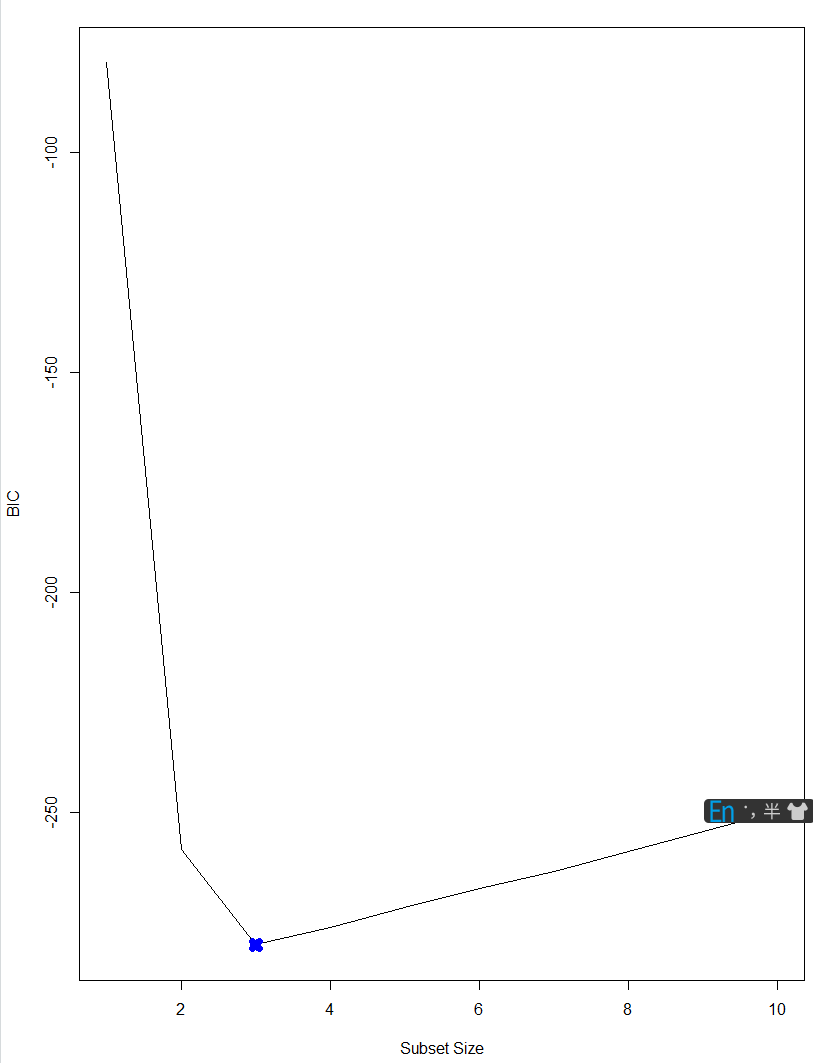
Pch=20表示小圆点，lwd是线条宽度

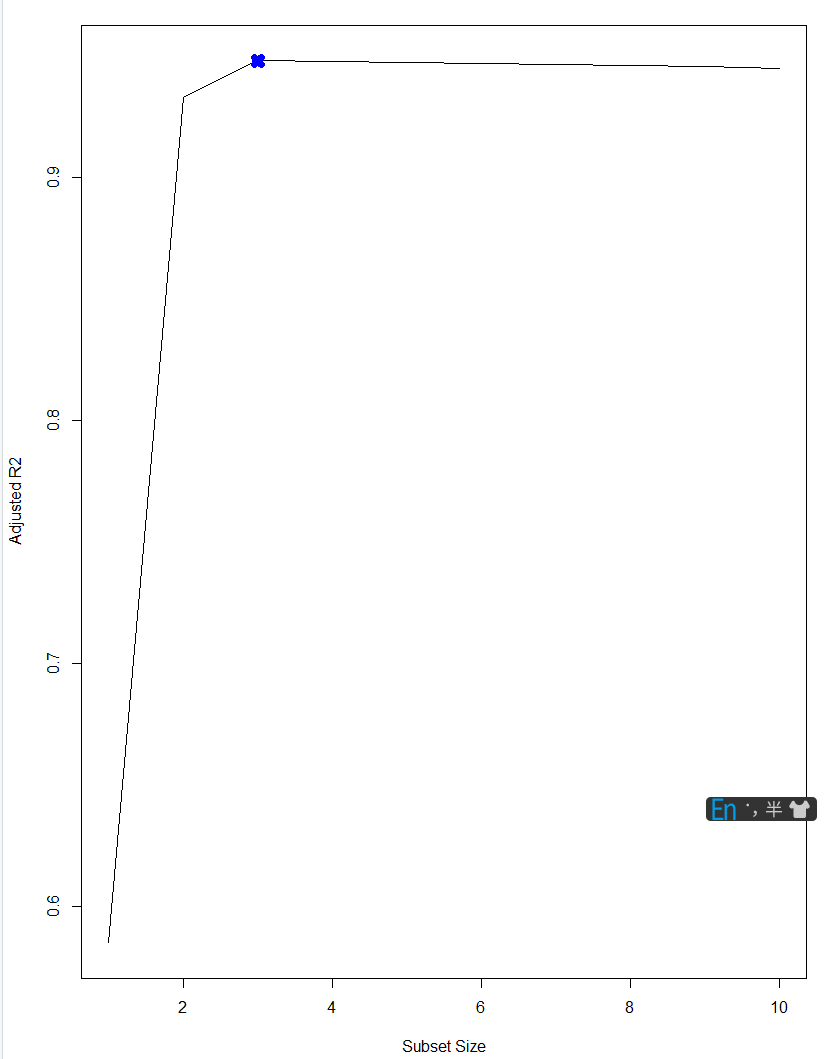
8（a-c）

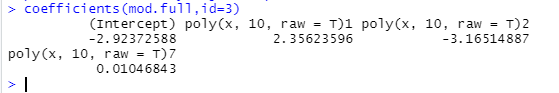


8.（c）





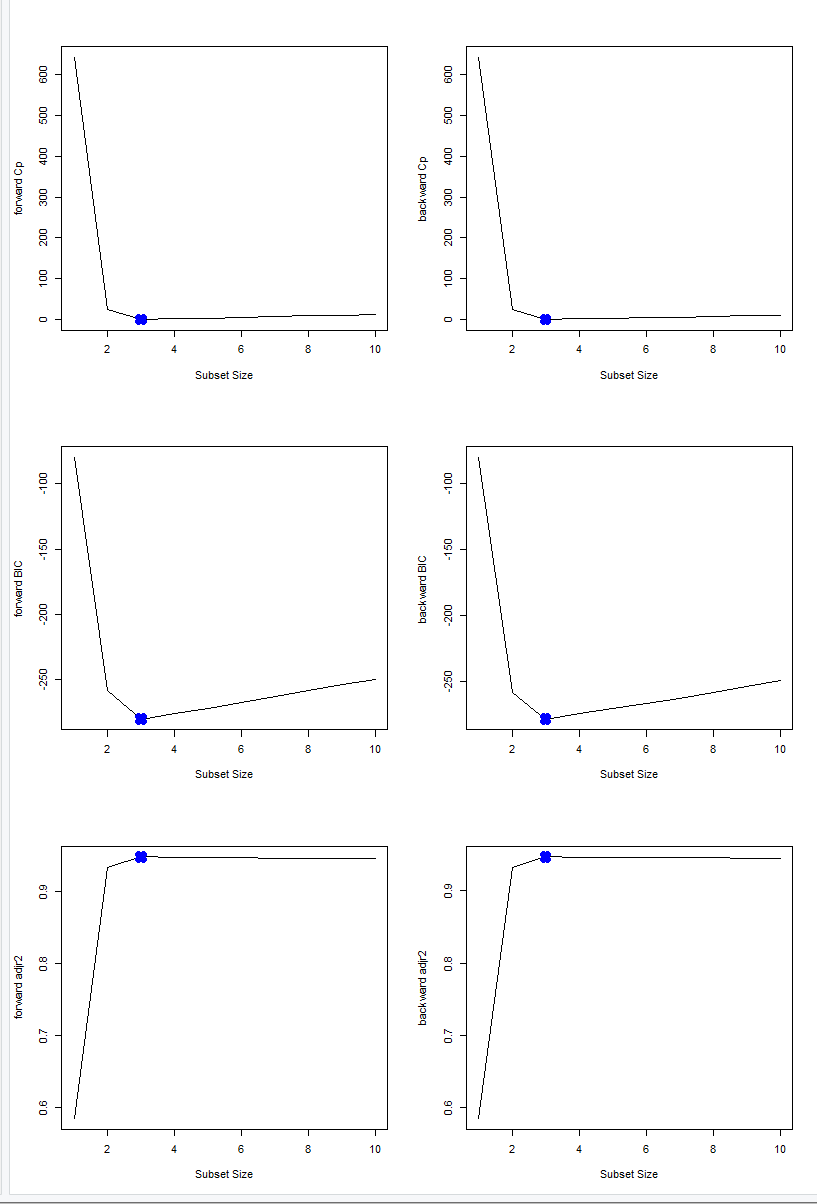


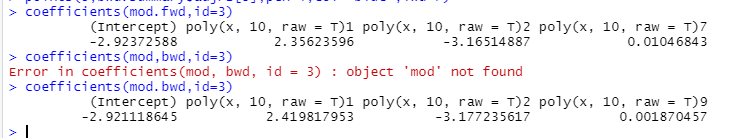


所以选择X7

8（d）

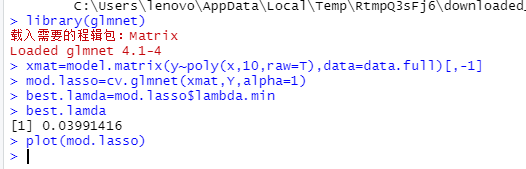


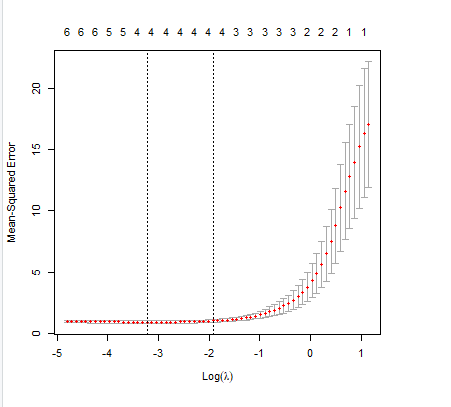


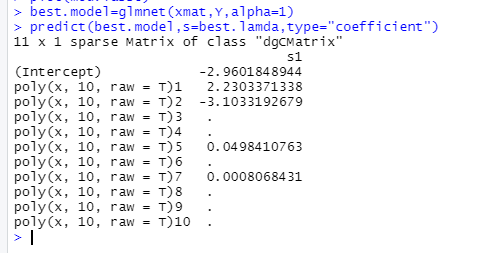


Forward选X7.backward选X9

8（e）

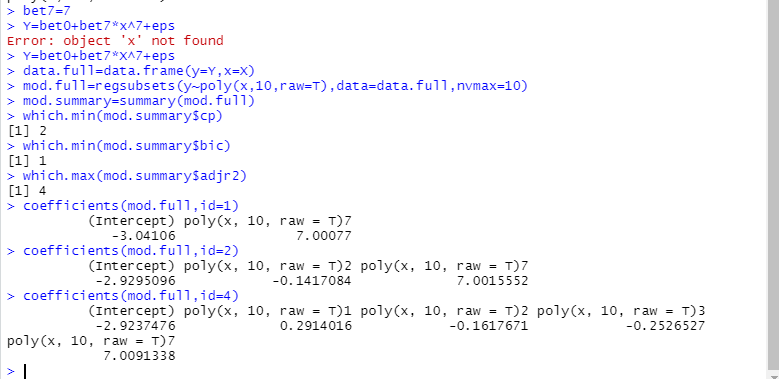




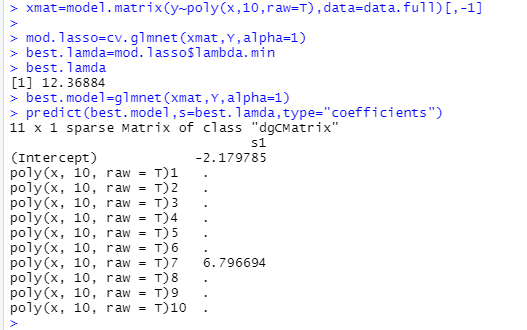


选择X5或者X7

8（f）

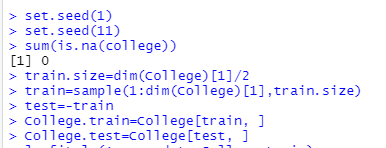


可以看到bic的系数最小

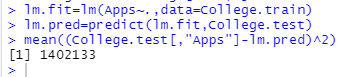


可以看到lasso也选择了最佳的单变量模型，但intercet却有很大的偏差

9（a）

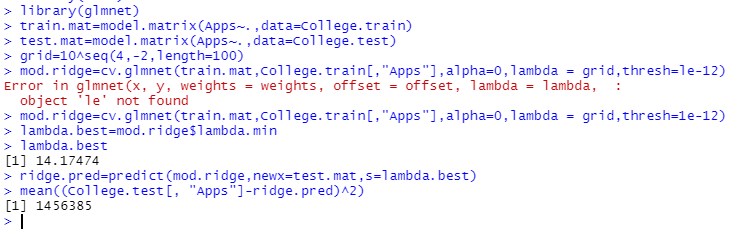


9（b）

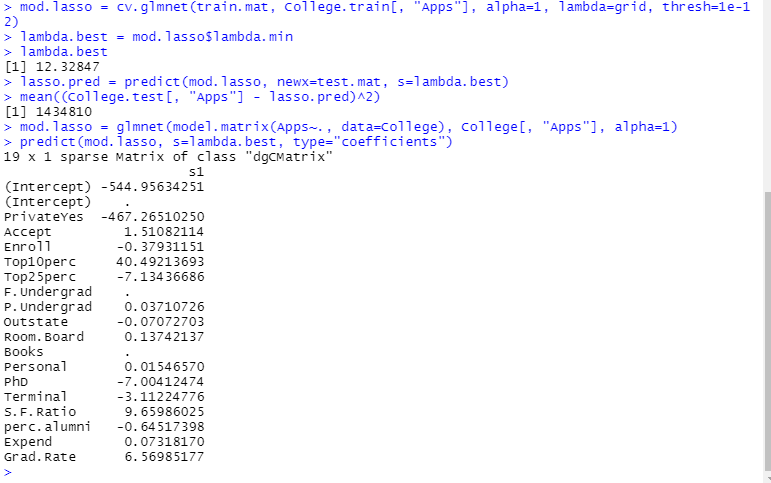


所以RSS为1402133

9（c）

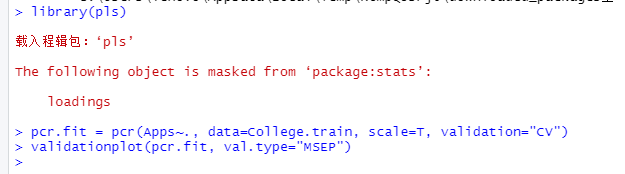


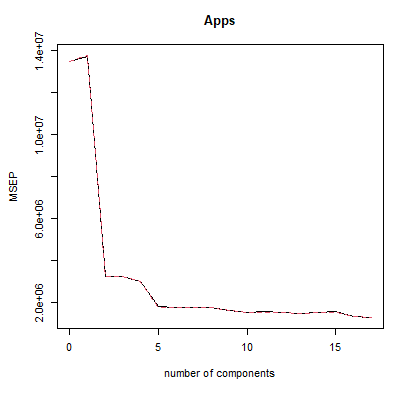
9（d）

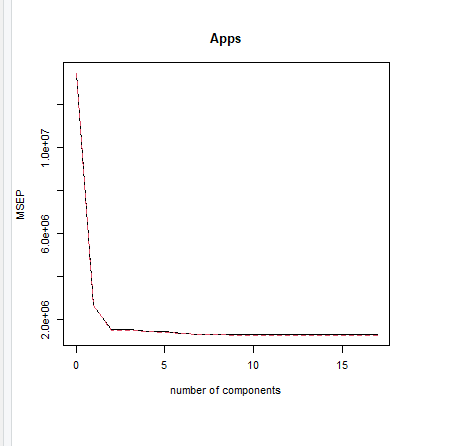


Mean算的是测试误差RSS

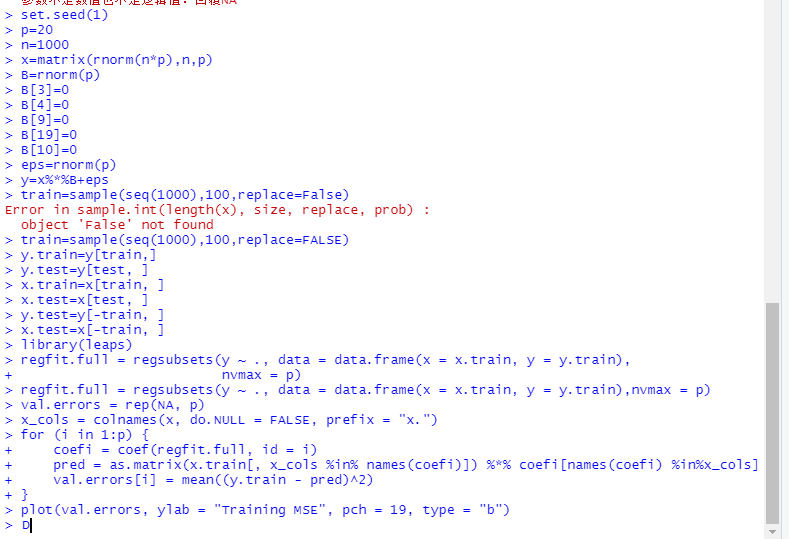
9（e）、

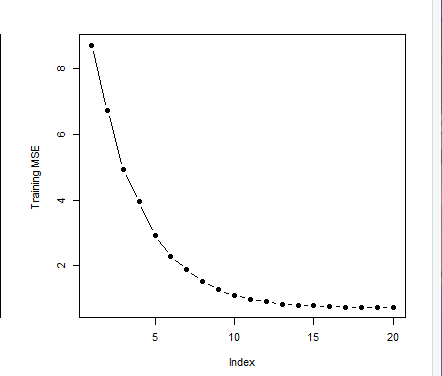




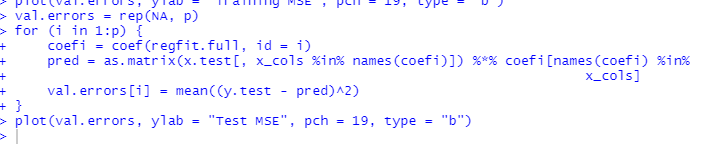


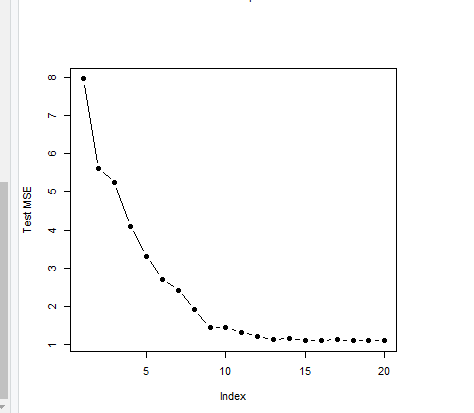
10（a-c）





10(d)

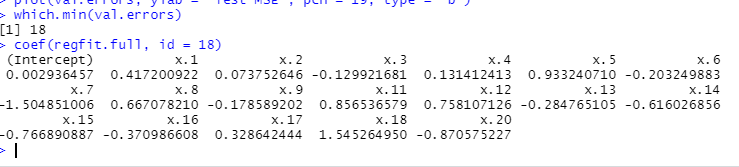




10(e)

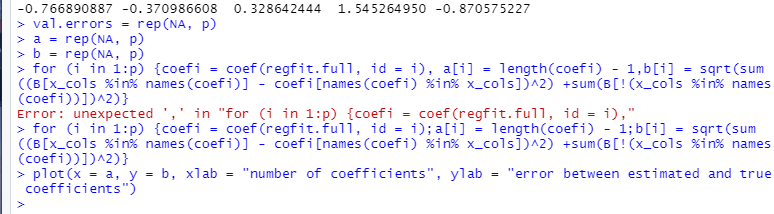


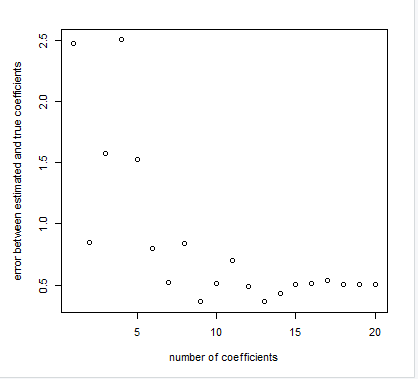
10(f)



除了x2有归零的系数外，其他的特征都被抓取了

10（g）







有9个系数的模型使估计值和真实系数之间的误差最小。9个参数的模型，测试误差达到最小。这里衡量的真实系数的拟合度较高，并不意味着该模型的测试MSE会更低。