Christina Wang

Econ 108

November 18, 2022

Problem Set 7

1)

a) Done

b) Done

c)

nearest5 = 0.9034783

nearest10 = 0.8826087

nearest30 = 0.8721739

nearest50 = 0.8695652

nearest100 = 0.8504348

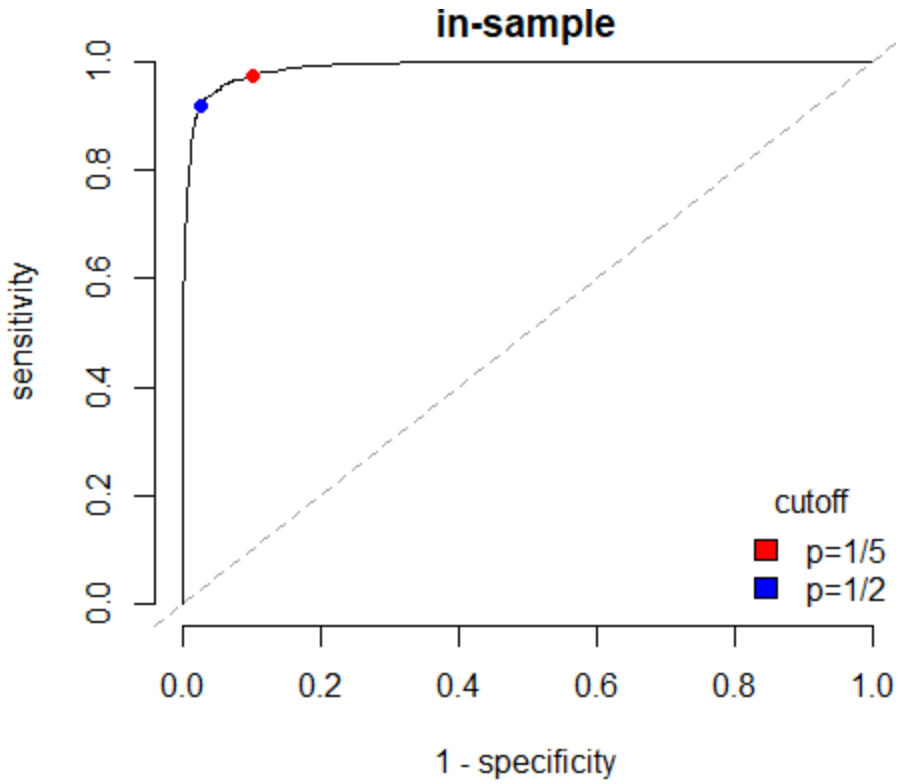
d)

K =5 is most accurate.

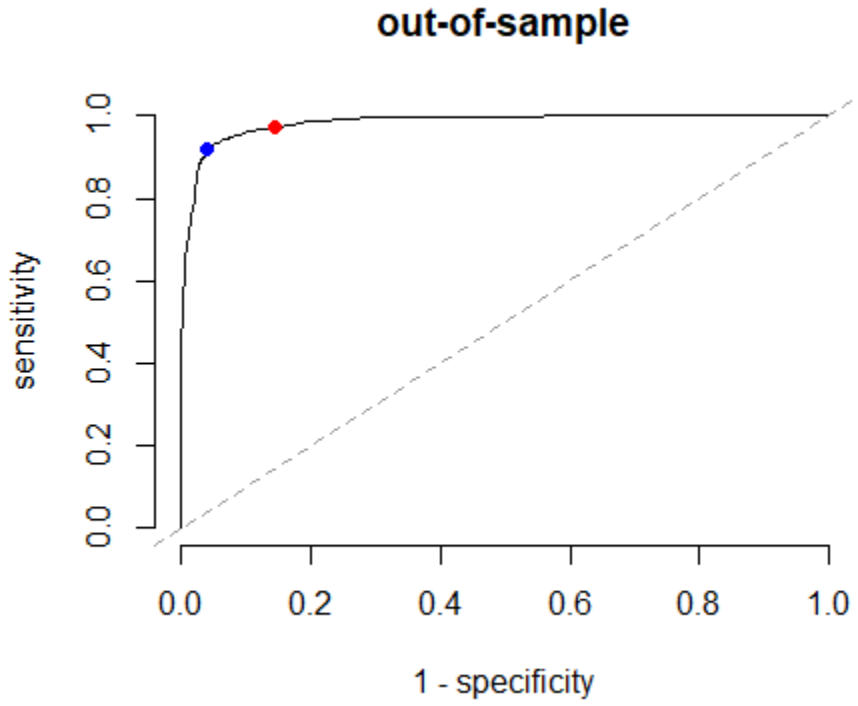
2)

a)

In Sample


in-sample
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
cutoff
p=1/5
p=1/2
0.0
0.2
0.4
0.6
0.8
1.0
1 - specificity

OOS

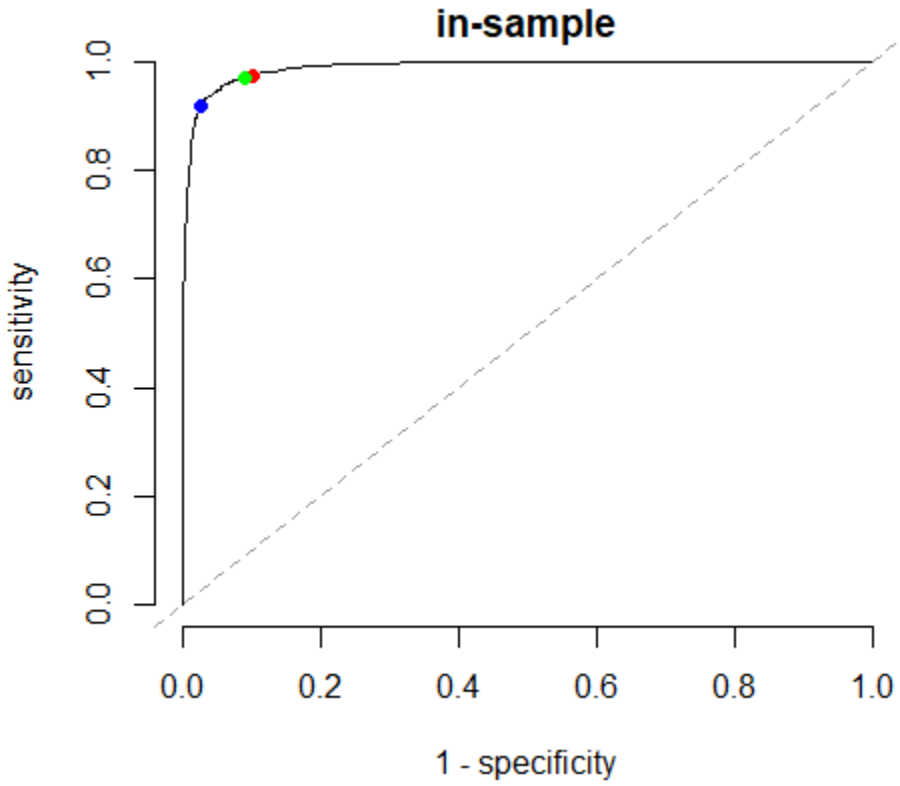

out-of-sample
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
1 - specificity

b)

c)

mean((pred>30/130)[Spam==1]) = .967

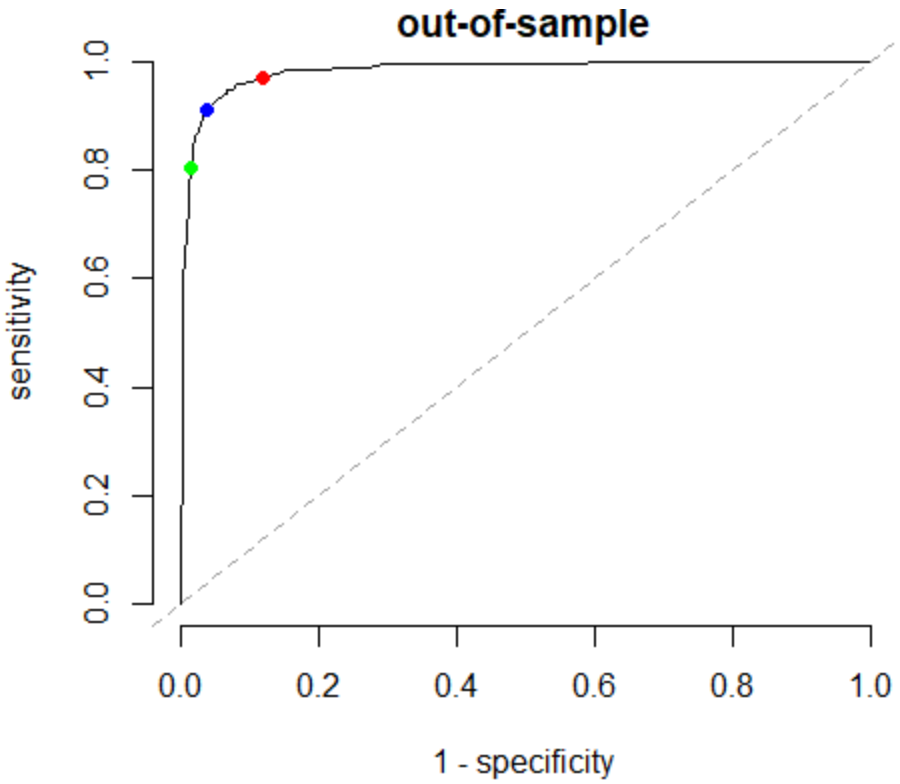
mean((pred<30/130)[Spam==0]) = .908

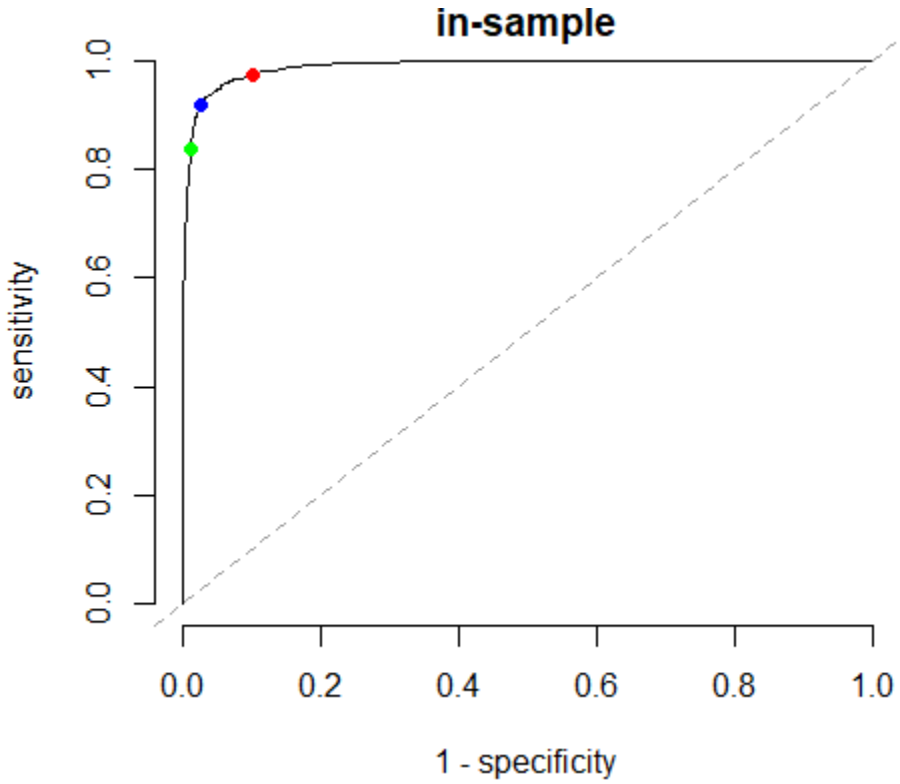


d)

mean((pred>100/130)[Spam==1])# sensitivity = .825

mean((pred<100/130)[Spam==0])# specificity = .987


out-of-sample
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - specificity


in-sample
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - specificity