Neha Maddali
ID: 110122037
Signature: Neha Maddali

Problem 1:

a)

	Best subset	forward	backward	Criteria value
M1	Y~X1	Y ~ X1	Y ~ X1	AIC
M2	Y~X2+X4	Y~X2+X4	Y~X2+X4	AIC
M3	Y~X1+X2+X3	Y~X1+X2+X3	Y~X1+X2+X3	BIC
M4	Y~X1+X2+X3+X4	Y~X1+X2+X3+X4	Y~X1+X2+X3+X4	AIC
Final model	Y ~ X1	Y ~ X1	Y ~ X1	Adjusted R^2

b) I think the student did a good job using the 10-fold CV to find the optimal tuning parameter. However, there is an alternative way. If a different training or test set or number of folds were selected for CV, the optimal λ and test error would also change. Instead, using the one-standard error rule can fix this issue. Instead of picking the λ that produces the smallest CV error, we pick the model whose CV error is within one standard error of the lowest point on the curve of the cross-validation error as a function of λ. This value would then be the λ^{ridge}_{1se}

c) False. When λ = 0, the bias is small but there will be a higher variance. When λ = ∞ the variance is small but the bias is high

d) True. Lasso is more flexible than least squares linear regression. Lasso will have a better prediction accuracy when its increase in variance is less than its decrease in bias

Problem 2:

a)

```
glm(formula = chd ~ age, family = "binomial", data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4321  -0.9215  -0.5392   1.0952   2.2433

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.521710   0.416031  -8.465  < 2e-16 ***
age          0.064108   0.008532   7.513 5.76e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 596.11  on 461  degrees of freedom
Residual deviance: 525.56  on 460  degrees of freedom
AIC: 529.56

Number of Fisher Scoring iterations: 4
```

The maximum likelihood estimate of b0 = -3.521710 and b1 = 0.064108
exp(beta0+beta1*462)/(1+exp(beta0+beta1*462)) = 1

b) Predicted probability = 0.4215756

c) standard error = 0

```
B = 2000
n=462
medhat = rep(NA,2000)
for(b in 1:B){
  index = sample(1:n,n,replace=TRUE)
  bootstrap = heart[index,]
  medhat[b] = median(bootstrap$chd, na.rm=TRUE)
}
sqrt(sum((medhat-mean(medhat))^2)/(B-1))
```

d) At the age of 55

Problem 3:

a)

```
       101        102        103        104        105 
0.61158280 0.11903050 0.17727017 0.08279173 0.17219104 
       106        107        108        109        110 
0.62235930 0.53401573 0.38208696 0.33339837 0.05302398 
```

First 10 predicted probabilities from training set

b) Threshold of 0.4 gives us a misclassification rate of 0.3181

```
glm.pred  0  1
     No   3  0
     Yes 14 27
```

c) Threshold of 0.5 gives us the smallest false positive rate. But the misclassification rate will then be 0.3421 which is larger than that of part b

```
glm.pred  0  1
     No   3  2
     Yes 11 22
```

d) An issue is that regularized regression is often used to control for effects of other predictors. In this scenario, least squares/regularized regression are not appropriate because they are not built for classifying data points. Logistic regression performs a better job at classifying these data points and has a better logarithmic loss function