

# logistic

Xinran Sun, Haotian Wu, Lin Yang, Shengzhi Luo

3/17/2022

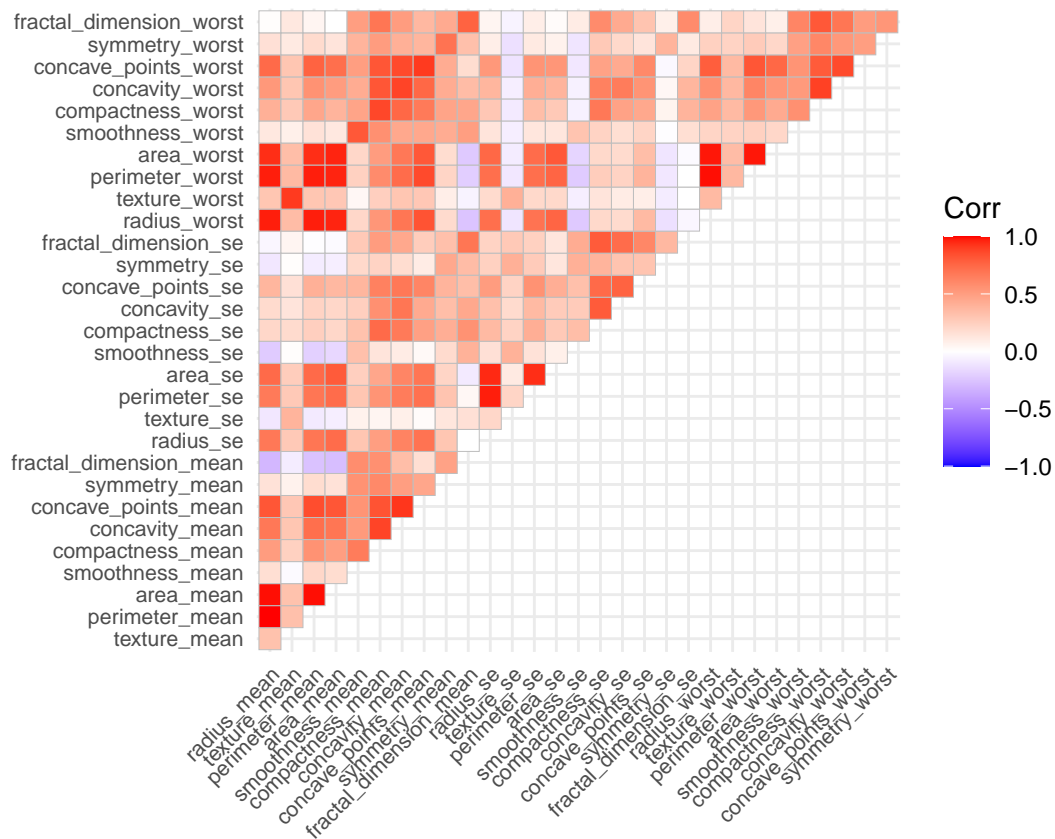
```
ggplot2::theme_set(theme_minimal() + theme(legend.position = "bottom"))
```

## data import and data clean

```
#load the data
breast = read.csv("breast-cancer.csv") %>%
  janitor::clean_names() %>%
  dplyr::select(-1, -33) %>% #drop id and NA columns
  mutate(diagnosis = recode(diagnosis, "M" = 1, "B" = 0))

#check collinearity
corr = breast[2:31] %>%
  cor()

ggcorrplot(corr, type = "upper", tl.cex = 8)
```

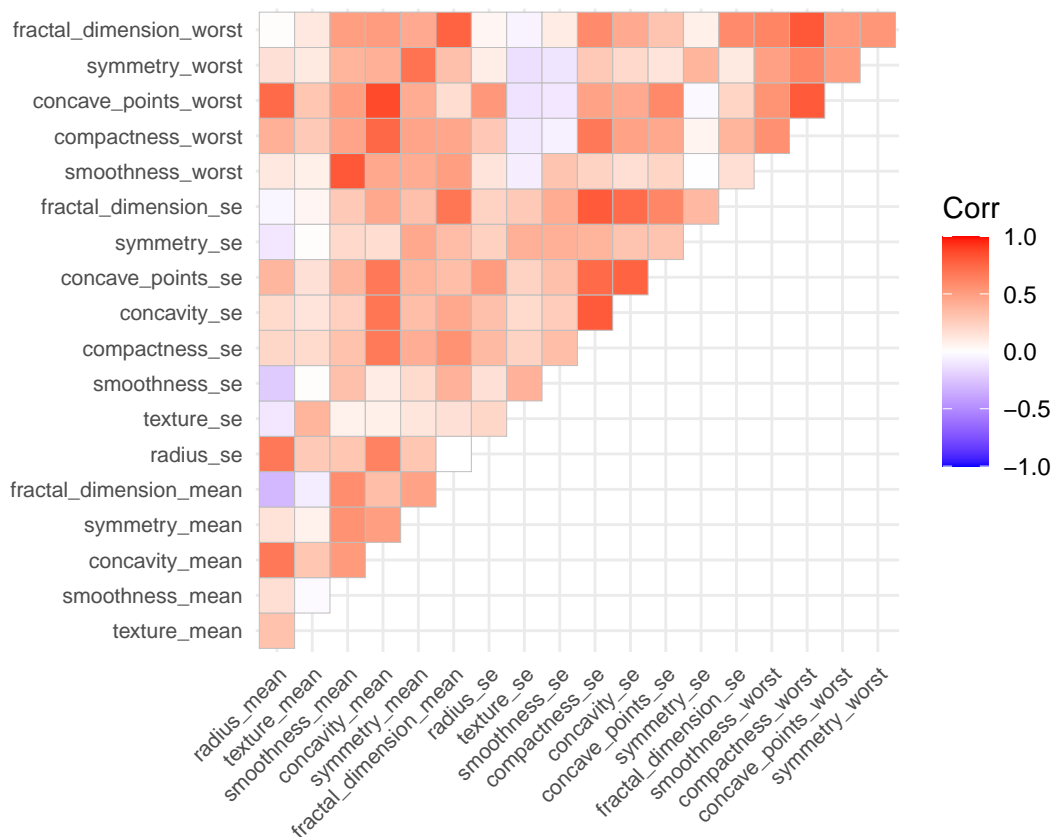


```
#remove some highly correlated variables
```

```
breast_dat <- breast %>% dplyr::select(-area_se, -perimeter_se, -area_worst, -perimeter_mean, -perimeter_worst)
```

```
corr1 = breast_dat[2:20] %>%  
  cor()
```

```
ggcorrplot(corr1, type = "upper", tl.cex = 8)
```



```
#partition data into training and test data
trainRows <- createDataPartition(y = breast_dat$diagnosis, p = 0.8, list = FALSE)
breast_train <- breast_dat[trainRows, ]
breast_test <- breast_dat[-trainRows, ]

head(breast_dat, 5)
```

```
##   diagnosis radius_mean texture_mean smoothness_mean concavity_mean
## 1         1      17.99      10.38      0.11840      0.3001
## 2         1      20.57      17.77      0.08474      0.0869
## 3         1      19.69      21.25      0.10960      0.1974
## 4         1      11.42      20.38      0.14250      0.2414
## 5         1      20.29      14.34      0.10030      0.1980
##   symmetry_mean fractal_dimension_mean radius_se texture_se smoothness_se
## 1      0.2419      0.07871      1.0950      0.9053      0.006399
## 2      0.1812      0.05667      0.5435      0.7339      0.005225
## 3      0.2069      0.05999      0.7456      0.7869      0.006150
## 4      0.2597      0.09744      0.4956      1.1560      0.009110
## 5      0.1809      0.05883      0.7572      0.7813      0.011490
##   compactness_se concavity_se concave_points_se symmetry_se
## 1      0.04904      0.05373      0.01587      0.03003
## 2      0.01308      0.01860      0.01340      0.01389
## 3      0.04006      0.03832      0.02058      0.02250
## 4      0.07458      0.05661      0.01867      0.05963
## 5      0.02461      0.05688      0.01885      0.01756
##   fractal_dimension_se smoothness_worst compactness_worst concave_points_worst
```

```
## 1      0.006193      0.1622      0.6656      0.2654
## 2      0.003532      0.1238      0.1866      0.1860
## 3      0.004571      0.1444      0.4245      0.2430
## 4      0.009208      0.2098      0.8663      0.2575
## 5      0.005115      0.1374      0.2050      0.1625
## symmetry_worst fractal_dimension_worst
## 1      0.4601      0.11890
## 2      0.2750      0.08902
## 3      0.3613      0.08758
## 4      0.6638      0.17300
## 5      0.2364      0.07678
```

```
r = dim(breast_dat)[1] #row number
c = dim(breast_dat)[2] #column number

var_names = names(breast_dat)[-c(1,2)] #variable names

standardize = function(col) {
  mean = mean(col)
  sd = sd(col)
  return((col - mean)/sd)
}

stand_df = breast_dat %>%
  dplyr::select(radius_mean:fractal_dimension_worst) %>%
  map_df(.x = ., standardize) #standardize

X = stand_df #predictors
y = as.vector(ifelse(breast_dat[,2] == "M", 1, 0)) #response
```

```
x_train <- breast_train[2:20] #predictors
y_train <- breast_train[1] #response
x_train_stan <- cbind(rep(1, nrow(x_train)), scale(x_train))

x_test <- breast_test[2:20]
x_test_stan <- cbind(rep(1, nrow(x_test)), scale(x_test))
```

## Full logistic model

```
glm.fit <- glm(diagnosis ~ .,
               data = breast_dat,
               subset = trainRows,
               family = binomial(link = "logit"))
summary(glm.fit)

##
## Call:
## glm(formula = diagnosis ~ ., family = binomial(link = "logit"),
##      data = breast_dat, subset = trainRows)
##
## Deviance Residuals:
```

```
##      Min      1Q      Median      3Q      Max
## -1.56866 -0.00003  0.00000  0.00000  2.82897
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -85.8416    36.2655  -2.367  0.0179 *
## radius_mean      0.4828     0.5136   0.940  0.3472
## texture_mean     1.2278     0.4834   2.540  0.0111 *
## smoothness_mean  -61.2848   138.3795  -0.443  0.6579
## concavity_mean   277.1480   130.0326   2.131  0.0331 *
## symmetry_mean   -174.8730    92.2200  -1.896  0.0579 .
## fractal_dimension_mean -268.5035  380.8394  -0.705  0.4808
## radius_se       68.8555    28.0557   2.454  0.0141 *
## texture_se      -0.4983     2.0041  -0.249  0.8036
## smoothness_se   1374.2231   645.1713   2.130  0.0332 *
## compactness_se   42.7437    215.7816   0.198  0.8430
## concavity_se    -150.6774    83.3050  -1.809  0.0705 .
## concave_points_se 731.2978   563.0794   1.299  0.1940
## symmetry_se    -1308.6896   653.0464  -2.004  0.0451 *
## fractal_dimension_se -6434.2765  2965.7057  -2.170  0.0300 *
## smoothness_worst -133.9879    98.5470  -1.360  0.1739
## compactness_worst -104.6221    49.1751  -2.128  0.0334 *
## concave_points_worst 161.9938    89.1095   1.818  0.0691 .
## symmetry_worst   180.8713    87.6875   2.063  0.0391 *
## fractal_dimension_worst 807.8439   356.3095   2.267  0.0234 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 599.11  on 455  degrees of freedom
## Residual deviance:  27.13  on 436  degrees of freedom
## AIC: 67.13
##
## Number of Fisher Scoring iterations: 13
```

```
pred <- predict(glm.fit, newdata = breast_test, type = "response")
y_test <- factor(breast_test$diagnosis)
auc_full <- auc(y_test, pred)
auc_full
```

```
## Area under the curve: 0.9771
```

## Newton-Raphson algorithm

### coordinate-wise optimization of a logistic-lasso model

```
#soft threshold
sfxn <- function(beta, lambda) {
  if (abs(beta) > lambda) {
    return(sign(beta) * (abs(beta) - lambda))
  }
}
```

```

}
else {
  return(0)
}
}

#coordinate-wise optimization function
coordwise_lasso <- function(lambda, x, y, betastart, tol = exp(-10), maxiter = 5000) {
  i <- 0
  n <- length(y)
  pnum <- length(betastart)
  betavec <- betastart
  loglik <- 0
  res <- c(0, loglik, betavec)
  prevloglik <- -Inf
  while (i < maxiter & abs(loglik - prevloglik) > tol & loglik < Inf) {
    i <- i + 1
    prevloglik <- loglik
    for (j in 1:pnum) {
      theta <- x %*% betavec
      p <- exp(theta) / (1 + exp(theta)) #probability of malignant cases
      w <- p*(1-p) #working weights
      w <- ifelse(abs(w-0) < 1e-5, 1e-5, w)
      z <- theta + (y - p)/w #working response
      zwoj <- x[, -j] %*% betavec[-j]
      betavec[j] <- sfxn(sum(w*(x[,j])*(z - zwoj)), lambda) / (sum(w*x[,j]*x[,j]))
    }
    theta <- x %*% betavec
    p <- exp(theta) / (1 + exp(theta)) #probability of malignant cases
    w <- p*(1-p) #working weights
    w <- ifelse(abs(w-0) < 1e-10, 1e-10, w)
    z <- theta + (y - p)/w
    loglik <- sum(w*(z - theta)^2) / (2*n) + lambda * sum(abs(betavec))
    res <- rbind(res, c(i, loglik, betavec))
  }
  return(res)
}

#coordwise_res <- coordwise_lasso(lambda = 0.006, x_train_stan, y_train, betastart = rep(0, #20))
#coordwise_res[nrow(coordwise_res), ]

```

We need to calculate lambdamax first to define a sequence of lambda.

```

x.matrix <- scale(x_train) %>% as.matrix()
y.matrix <- as.matrix(y_train)
lambdamax <- max(abs(t(x.matrix) %*% y.matrix)) #/ nrow(y.matrix)
lambda_seq1 <- exp(seq(log(lambdamax), -5, length = 50))
lambda_seq2 <- exp(seq(log(lambdamax), -5, length = 50))

#a path of solutions
pathwise <- function(x, y, lambda) {
  n <- length(lambda)
  betastart <- rep(0, 20)
  betas <- NULL

```



lambda	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21
2.2465820	2.01594281	1.46200675	2.19000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.43927020	2.00089468	16.8290000		
0.6396141											0.4290305		0.2614635	2.6765						
1.8256408	2.07952329	5.90000005	8.99300000	0.02866137	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.46423240	2.00041765	5.4080000		
0.6318917											0.5274154		0.3850862	1.4642						
1.4835711	2.13623776	1.09500065	4.69200000	2.50459890	0.000093195	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.48403000	2.00381562	8.9000000		
0.6247137						0.0094589					0.6343701		0.5100918	8.8043						
1.2055949	2.11003932	2.00000481	4.125	0.00000701	4.80000026	2.341	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.39307850	2.00453958	5.3000000		
0.6204545						0.0334200					0.8354201		0.6937958	2.3518						
0.9797031	2.07164684	1.00200801	1.700	0.00000009	2.45800023	5.351	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.31795850	2.00737167	9.0250000		
0.6105199						0.1273622					1.0432926		0.8585878	1.9011						
0.7961365	2.05835217	3.00400067	8.661	-	3.45227941	0.000378	3.609	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.18738580	2.00727.2226	3.738494		
0.5713337						0.19220927	2.910				1.3045021		0.9976482	8.7000						
0.6469647	2.03367170	1.20800507	2.363	-	3.80948000	0.00521	3.532	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.06282450	2.00711.2040	4.0656093		
0.5263515						0.2512397	5.9713				1.5780634		1.1253283	3.7735						
0.5257431	2.05017947	1.56500801	1.989	-	4.16798300	0.0062	6.9577	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.00000000	2.00576.7448	5.835882		
0.4818089						0.32100505	1.391				1.8072508		1.2366239	1.1216						
0.4272348	2.10028778	1.40000020	5.446	-	4.52408310	0.0060	8.912	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	-	0.00000000	2.00720.5233	7.751544		
0.4330860						0.39702208	2.752				2.0154409		1.3189927	2.9992						
0.3471840	2.12648897	1.25700501	3.429	-	4.88280490	0.0070	4.1906	-	0.00000000	-	0.00000000	-	0.00000000	-	2.8631679	1.7000	9.793			
0.3821628						0.4970501	3.5814				2.0550745	0.8868	1.4446735	2.6781	0.2833424					
0.2821323	2.11253237	1.48700055	0.954	-	5.25579550	0.0070	4.4717	-	0.00000000	-	0.00000000	-	0.00000000	-	3.08528289	4.526585				
0.3243007						0.62920821	5.336				1.9110437	2.406	1.6259258	1.5449	0.8669767					
0.2292692	2.02280148	2.00400012	5.630	-	5.67263500	0.0080	7.2601	-	0.00000000	-	0.00000000	-	0.00000000	-	3.40132822	9.400293				
0.2789284						0.7560403	4.9233				1.7932732	4.1607	1.8891094	8.3557	1.44811362					
0.1863111	1.89121026	1.67600027	3.6115	-	6.09929730	0.0008	3.4611	-	0.00000000	-	0.00000000	-	0.00000000	-	3.77825892	6.580328				
0.2516542						0.9019818	3.3770				1.5579535	8.8667	2.1982958	9.8395	2.57202460					
0.1514021	1.75826841	1.50000009	2.105	-	6.55829940	0.0008	3.8451	-	0.00000000	-	0.00000000	-	0.00000000	-	4.18925230	2.377305				
0.2350067						1.0618870	1.07694				1.2792373	2.5245	2.5343926	0.0323	6.379589577					
0.1230339	1.65211335	8.469	6.8421318	-	7.0694318	1.4917	1.87	-	0.2013255	-	0.00000000	-	0.00000000	-	4.45130336	2.183653				
0.2436290						1.2350967	7.9707	0.0282237	1.0660178	8.2783	2.9334564	9.0987	3.27821	4.850						
0.0999811	1.57224853	1.112	7.7114249	-	7.6280353	1.6718	8.72	-	0.4802175	-	0.00000000	-	0.00000000	-	4.68136774	2.2827317				
0.2632320						1.1196103	1.414068	6.2344	0.0624598	0.935672	4.7815	3.3562957	4.7940	1.4834321						
0.0812477	1.5841513	2.0271	7.8581896	-	7.8442642	1.6917	6.11	-	0.5535705	-	0.00000000	-	0.00000000	-	4.8213811	1.8886	7.455			
0.2704924						0.1502201	1.485008	7.6607	0.0633106	1.03015	7.8553	6.7	3.431119	4.3847	6.945691522					
0.0660243	1.4847782	4.2797	9.2211121	-	8.7177878	2.0153	0.37	-	0.8918518	-	0.00000000	-	0.00000000	-	5.22808253	6.282987				
0.3012377						0.2562806	1.747953	3.93630	0.1016379	0.924784	3.2698	4.104265	2.7373	1.216668	3.7296					
0.0536534	1.4662595	5.898	10.5188493	-	9.5640994	2.2571	7.79	-	1.5149891	-	0.00000000	-	0.00000000	-	5.51833261	8.146122				
0.4077028						0.3606266	2.053630	7.70349	0.1589325	0.2962100	5.8805	4.800668	2.0795	4.927281	7.2269					
0.0436004	1.4647369	5.232	11.4471155	-	10.2766266	2.4556	6.77	-	1.8221497	-	0.00000000	-	0.00000000	-	5.89473334	6.0962438				
0.4585712						0.4194876	2.284590	2.13792	0.1819305	0.125836	4.96368	5.292858	4.1283	2.6785	1.6074					
0.0354310	1.4730326	3.786	12.3706586	-	10.9925132	2.6457	4.0200000	2.1116853	-	-	0.00000000	-	0.00000000	-	6.28813138	4.032452				
0.5058222						0.4729787	2.515371	5.86186	0.2007129		2.5883931	5.780918	6.0522	5.8650	9.484					
0.0287923	1.4976532	3.913	13.2834056	-	11.6877519	2.8279	2.0000000	2.3759244	-	-	0.00000000	-	0.00000000	-	6.6716872	7.0810521				
0.5449859						0.5239582	2.735662	4.1103	0.2108514		2.7911489	6.25293	5.1864	9.9087	5.4568					
0.0233975	1.5221467	2.7267	14.1759302	-	12.3653977	3.0005	1.0000000	2.6241298	-	-	0.00000000	-	0.00000000	-	7.05659576	4.319670				
0.5820266						0.5673982	2.950820	3.7531	0.2189496		2.9865965	6.7163	7.3210	7.2812	2.19615					
0.0190135	1.5102436	8.477	14.2492477	-	12.4767243	3.0182	6.0200000	2.6287574	-	-	0.00000000	-	0.00000000	-	7.15844538	2.730981				
0.5876231						0.5733327	2.985651	6.3809	0.2238884		3.0019211	6.7651	3.3282	3.2482	5.3982752					
0.0154510	1.5608208	6.737	15.8530908	-	13.638630	3.3117	3.2017067	3.0818013	-	-	0.00000000	-	0.00000000	-	7.8109795	7.7697713				
0.6547548						0.6348092	3.360340	2.76422	0.2349384		3.3707509	7.5992	2.977	2.7256	5.0535	3.5923				
0.0125559	1.571488	7.2212	16.6117316	-	14.219792	6.3447	1.0283	4.71	3.2923269	-	-	0.00000000	-	0.00000000	-	8.1689472	9.8396171			
0.6915566						0.6617006	3.550750	7.60908	0.2451917		3.5616824	8.0061	3.2824	4.7882	9.084858					



lambda	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21
0.0102033	1.5826207	1.590993	17.3113466	-	14.75091253	5.705365	1430	3.4837598	-	-	-	8.49783700	36.5963981							
	0.7248177		0.6859605	3.7248759	86784	0.2534951		3.7342582	8.3804411	944253016	014759									
0.0082915	1.5946205	1.590775	17.9474420	-	15.22983663	6.812385	1277	3.6558746	-	-	-	8.7943113	8207050							
	0.7545704		0.7076959	3.8820037	48911	0.2601638		3.8885088	8.7201285	96504781	5324081									
0.0067379	1.5938426	1.591886	17.9646426	-	15.26087743	6.847673	2585	3.6591624	-	-	-	8.8224644	7982376605							
	0.7570750		0.7111196	3.8912782	70599	0.2627046		3.8906583	8.7313757	006952861	2453739									

## cross-validation

```
set.seed(2022)

cv = function(data, lambda) {
  n <- nrow(data)
  data <- data[sample(n), ] #shuffle the data
  folds <- cut(seq(1, nrow(data)), breaks = 5, labels = FALSE) #Create 5 equal size folds
  # mse <- data.frame() #a data frame storing mse results
  #mse_lambda <- vector()
  #se <- vector() #a vector storing test errors
  res <- lambda
  #se <- vector() #a vectro storing test errors

  #Perform 5 fold cross validation
  for (i in 1:5) {
    #partition the data into train and test data
    testRows <- which(folds == i, arr.ind = TRUE)
    data_test <- data[testRows, ]
    data_train <- data[-testRows, ]
    x_train <- data_train[2:20]
    x_train_stan <- cbind(rep(1, nrow(x_train)), scale(x_train))
    y_train <- data_train[1]
    x_test <- data_test[2:20]
    #standardized test data
    x_test_stan <- cbind(rep(1, nrow(x_test)), scale(x_test))
    y_test <- data_test %>% mutate(diagnosis = factor(diagnosis))
    y_test <- y_test$diagnosis
    #Use the test and train data partitions to perform lasso
    path_sol <- pathwise(x = x_train_stan,
                        y = y_train,
                        lambda = lambda)

    auc <- vector()
    for (j in 1:length(lambda)) {
      curbeta <- as.numeric(path_sol[j, 2:21])
      theta <- x_test_stan %*% curbeta
      p <- exp(theta) / (1 + exp(theta))
      auc[j] <- auc(y_test, p)
      #y.pred <- ifelse(p > 0.5, 1, 0)
      #accuracy[j] <- mean(y.pred == y_test)
    }
    print(auc)
    res <- cbind(res, auc)
  }
}
```

```

    print(res)
  }
  return(res)
  #se[j] <- sqrt(var(error)/5)
  #cv.auc.lambda <- rowMeans(mse)
  #return(cv.auc.lambda)
}

cv_test = cv(data = breast_train, lambda_seq2)

```

```

## [1] 0.5000000 0.5000000 0.9931253 0.9931253 0.9931253 0.9920677 0.9910100
## [8] 0.9899524 0.9899524 0.9910100 0.9915389 0.9920677 0.9931253 0.9952406
## [15] 0.9957694 0.9962983 0.9968271 0.9968271 0.9968271 0.9968271 0.9968271
## [22] 0.9968271 0.9962983 0.9962983 0.9957694 0.9957694 0.9957694 0.9952406
## [29] 0.9957694 0.9952406 0.9931253 0.9925965 0.9925965 0.9931253 0.9931253
## [36] 0.9936542 0.9931253 0.9920677 0.9920677 0.9915389 0.9894236 0.9894236
## [43] 0.9894236 0.9894236 0.9746166 0.9746166 0.9746166 0.9746166 0.9746166
## [50] 0.9746166
##           res           auc
## [1,] 1.752945e+02 0.5000000
## [2,] 1.424496e+02 0.5000000
## [3,] 1.157589e+02 0.9931253
## [4,] 9.406920e+01 0.9931253
## [5,] 7.644348e+01 0.9931253
## [6,] 6.212030e+01 0.9920677
## [7,] 5.048084e+01 0.9910100
## [8,] 4.102226e+01 0.9899524
## [9,] 3.333594e+01 0.9899524
## [10,] 2.708979e+01 0.9910100
## [11,] 2.201399e+01 0.9915389
## [12,] 1.788924e+01 0.9920677
## [13,] 1.453734e+01 0.9931253
## [14,] 1.181348e+01 0.9952406
## [15,] 9.599994e+00 0.9957694
## [16,] 7.801246e+00 0.9962983
## [17,] 6.339530e+00 0.9968271
## [18,] 5.151694e+00 0.9968271
## [19,] 4.186423e+00 0.9968271
## [20,] 3.402015e+00 0.9968271
## [21,] 2.764580e+00 0.9968271
## [22,] 2.246582e+00 0.9968271
## [23,] 1.825641e+00 0.9962983
## [24,] 1.483571e+00 0.9962983
## [25,] 1.205595e+00 0.9957694
## [26,] 9.797031e-01 0.9957694
## [27,] 7.961365e-01 0.9957694
## [28,] 6.469647e-01 0.9952406
## [29,] 5.257431e-01 0.9957694
## [30,] 4.272348e-01 0.9952406
## [31,] 3.471840e-01 0.9931253
## [32,] 2.821323e-01 0.9925965
## [33,] 2.292692e-01 0.9925965
## [34,] 1.863111e-01 0.9931253

```

```

## [35,] 1.514021e-01 0.9931253
## [36,] 1.230339e-01 0.9936542
## [37,] 9.998112e-02 0.9931253
## [38,] 8.124769e-02 0.9920677
## [39,] 6.602434e-02 0.9920677
## [40,] 5.365338e-02 0.9915389
## [41,] 4.360037e-02 0.9894236
## [42,] 3.543098e-02 0.9894236
## [43,] 2.879229e-02 0.9894236
## [44,] 2.339749e-02 0.9894236
## [45,] 1.901351e-02 0.9746166
## [46,] 1.545096e-02 0.9746166
## [47,] 1.255592e-02 0.9746166
## [48,] 1.020332e-02 0.9746166
## [49,] 8.291527e-03 0.9746166
## [50,] 6.737947e-03 0.9746166
## [1] 0.5000000 0.5000000 0.9456989 0.9532258 0.9715054 0.9768817 0.9790323
## [8] 0.9795699 0.9806452 0.9860215 0.9876344 0.9897849 0.9919355 0.9930108
## [15] 0.9935484 0.9940860 0.9940860 0.9946237 0.9946237 0.9962366 0.9962366
## [22] 0.9962366 0.9962366 0.9962366 0.9962366 0.9978495 0.9978495 0.9973118
## [29] 0.9978495 0.9973118 0.9967742 0.9967742 0.9951613 0.9946237 0.9935484
## [36] 0.9935484 0.9935484 0.9935484 0.9919355 0.9924731 0.9919355 0.9919355
## [43] 0.9919355 0.9913978 0.9913978 0.9913978 0.9913978 0.9913978 0.9913978
## [50] 0.9919355
##           res           auc           auc
## [1,] 1.752945e+02 0.5000000 0.5000000
## [2,] 1.424496e+02 0.5000000 0.5000000
## [3,] 1.157589e+02 0.9931253 0.9456989
## [4,] 9.406920e+01 0.9931253 0.9532258
## [5,] 7.644348e+01 0.9931253 0.9715054
## [6,] 6.212030e+01 0.9920677 0.9768817
## [7,] 5.048084e+01 0.9910100 0.9790323
## [8,] 4.102226e+01 0.9899524 0.9795699
## [9,] 3.333594e+01 0.9899524 0.9806452
## [10,] 2.708979e+01 0.9910100 0.9860215
## [11,] 2.201399e+01 0.9915389 0.9876344
## [12,] 1.788924e+01 0.9920677 0.9897849
## [13,] 1.453734e+01 0.9931253 0.9919355
## [14,] 1.181348e+01 0.9952406 0.9930108
## [15,] 9.599994e+00 0.9957694 0.9935484
## [16,] 7.801246e+00 0.9962983 0.9940860
## [17,] 6.339530e+00 0.9968271 0.9940860
## [18,] 5.151694e+00 0.9968271 0.9946237
## [19,] 4.186423e+00 0.9968271 0.9946237
## [20,] 3.402015e+00 0.9968271 0.9962366
## [21,] 2.764580e+00 0.9968271 0.9962366
## [22,] 2.246582e+00 0.9968271 0.9962366
## [23,] 1.825641e+00 0.9962983 0.9962366
## [24,] 1.483571e+00 0.9962983 0.9962366
## [25,] 1.205595e+00 0.9957694 0.9962366
## [26,] 9.797031e-01 0.9957694 0.9978495
## [27,] 7.961365e-01 0.9957694 0.9978495
## [28,] 6.469647e-01 0.9952406 0.9973118
## [29,] 5.257431e-01 0.9957694 0.9978495

```

```

## [30,] 4.272348e-01 0.9952406 0.9973118
## [31,] 3.471840e-01 0.9931253 0.9967742
## [32,] 2.821323e-01 0.9925965 0.9967742
## [33,] 2.292692e-01 0.9925965 0.9951613
## [34,] 1.863111e-01 0.9931253 0.9946237
## [35,] 1.514021e-01 0.9931253 0.9935484
## [36,] 1.230339e-01 0.9936542 0.9935484
## [37,] 9.998112e-02 0.9931253 0.9935484
## [38,] 8.124769e-02 0.9920677 0.9935484
## [39,] 6.602434e-02 0.9920677 0.9919355
## [40,] 5.365338e-02 0.9915389 0.9924731
## [41,] 4.360037e-02 0.9894236 0.9919355
## [42,] 3.543098e-02 0.9894236 0.9919355
## [43,] 2.879229e-02 0.9894236 0.9919355
## [44,] 2.339749e-02 0.9894236 0.9913978
## [45,] 1.901351e-02 0.9746166 0.9913978
## [46,] 1.545096e-02 0.9746166 0.9913978
## [47,] 1.255592e-02 0.9746166 0.9913978
## [48,] 1.020332e-02 0.9746166 0.9913978
## [49,] 8.291527e-03 0.9746166 0.9919355
## [50,] 6.737947e-03 0.9746166 0.9919355
## [1] 0.5000000 0.5000000 0.9474747 0.9540404 0.9585859 0.9590909 0.9631313
## [8] 0.9636364 0.9641414 0.9681818 0.9702020 0.9742424 0.9757576 0.9767677
## [15] 0.9813131 0.9833333 0.9858586 0.9873737 0.9883838 0.9898990 0.9909091
## [22] 0.9904040 0.9898990 0.9868687 0.9858586 0.9848485 0.9843434 0.9833333
## [29] 0.9823232 0.9813131 0.9787879 0.9772727 0.9757576 0.9737374 0.9712121
## [36] 0.9696970 0.9681818 0.9666667 0.9676768 0.9676768 0.9671717 0.9671717
## [43] 0.9671717 0.9661616 0.9661616 0.9661616 0.9656566 0.9656566 0.9661616
## [50] 0.9661616
##          res          auc          auc          auc
## [1,] 1.752945e+02 0.5000000 0.5000000 0.5000000
## [2,] 1.424496e+02 0.5000000 0.5000000 0.5000000
## [3,] 1.157589e+02 0.9931253 0.9456989 0.9474747
## [4,] 9.406920e+01 0.9931253 0.9532258 0.9540404
## [5,] 7.644348e+01 0.9931253 0.9715054 0.9585859
## [6,] 6.212030e+01 0.9920677 0.9768817 0.9590909
## [7,] 5.048084e+01 0.9910100 0.9790323 0.9631313
## [8,] 4.102226e+01 0.9899524 0.9795699 0.9636364
## [9,] 3.333594e+01 0.9899524 0.9806452 0.9641414
## [10,] 2.708979e+01 0.9910100 0.9860215 0.9681818
## [11,] 2.201399e+01 0.9915389 0.9876344 0.9702020
## [12,] 1.788924e+01 0.9920677 0.9897849 0.9742424
## [13,] 1.453734e+01 0.9931253 0.9919355 0.9757576
## [14,] 1.181348e+01 0.9952406 0.9930108 0.9767677
## [15,] 9.599994e+00 0.9957694 0.9935484 0.9813131
## [16,] 7.801246e+00 0.9962983 0.9940860 0.9833333
## [17,] 6.339530e+00 0.9968271 0.9940860 0.9858586
## [18,] 5.151694e+00 0.9968271 0.9946237 0.9873737
## [19,] 4.186423e+00 0.9968271 0.9946237 0.9883838
## [20,] 3.402015e+00 0.9968271 0.9962366 0.9898990
## [21,] 2.764580e+00 0.9968271 0.9962366 0.9909091
## [22,] 2.246582e+00 0.9968271 0.9962366 0.9904040
## [23,] 1.825641e+00 0.9962983 0.9962366 0.9898990
## [24,] 1.483571e+00 0.9962983 0.9962366 0.9868687

```

```

## [25,] 1.205595e+00 0.9957694 0.9962366 0.9858586
## [26,] 9.797031e-01 0.9957694 0.9978495 0.9848485
## [27,] 7.961365e-01 0.9957694 0.9978495 0.9843434
## [28,] 6.469647e-01 0.9952406 0.9973118 0.9833333
## [29,] 5.257431e-01 0.9957694 0.9978495 0.9823232
## [30,] 4.272348e-01 0.9952406 0.9973118 0.9813131
## [31,] 3.471840e-01 0.9931253 0.9967742 0.9787879
## [32,] 2.821323e-01 0.9925965 0.9967742 0.9772727
## [33,] 2.292692e-01 0.9925965 0.9951613 0.9757576
## [34,] 1.863111e-01 0.9931253 0.9946237 0.9737374
## [35,] 1.514021e-01 0.9931253 0.9935484 0.9712121
## [36,] 1.230339e-01 0.9936542 0.9935484 0.9696970
## [37,] 9.998112e-02 0.9931253 0.9935484 0.9681818
## [38,] 8.124769e-02 0.9920677 0.9935484 0.9666667
## [39,] 6.602434e-02 0.9920677 0.9919355 0.9676768
## [40,] 5.365338e-02 0.9915389 0.9924731 0.9676768
## [41,] 4.360037e-02 0.9894236 0.9919355 0.9671717
## [42,] 3.543098e-02 0.9894236 0.9919355 0.9671717
## [43,] 2.879229e-02 0.9894236 0.9919355 0.9671717
## [44,] 2.339749e-02 0.9894236 0.9913978 0.9661616
## [45,] 1.901351e-02 0.9746166 0.9913978 0.9661616
## [46,] 1.545096e-02 0.9746166 0.9913978 0.9661616
## [47,] 1.255592e-02 0.9746166 0.9913978 0.9656566
## [48,] 1.020332e-02 0.9746166 0.9913978 0.9656566
## [49,] 8.291527e-03 0.9746166 0.9919355 0.9661616
## [50,] 6.737947e-03 0.9746166 0.9919355 0.9661616
## [1] 0.5000000 0.5000000 0.9700742 0.9708686 0.9772246 0.9788136 0.9777542
## [8] 0.9782839 0.9804025 0.9830508 0.9819915 0.9814619 0.9804025 0.9804025
## [15] 0.9798729 0.9793432 0.9793432 0.9793432 0.9798729 0.9804025 0.9804025
## [22] 0.9804025 0.9825212 0.9830508 0.9846398 0.9856992 0.9867585 0.9878178
## [29] 0.9878178 0.9883475 0.9888771 0.9894068 0.9894068 0.9904661 0.9909958
## [36] 0.9904661 0.9909958 0.9899364 0.9899364 0.9920551 0.9931144 0.9931144
## [43] 0.9936441 0.9931144 0.9931144 0.9936441 0.9936441 0.9936441 0.9936441
## [50] 0.9936441
##          res          auc          auc          auc          auc
## [1,] 1.752945e+02 0.5000000 0.5000000 0.5000000 0.5000000
## [2,] 1.424496e+02 0.5000000 0.5000000 0.5000000 0.5000000
## [3,] 1.157589e+02 0.9931253 0.9456989 0.9474747 0.9700742
## [4,] 9.406920e+01 0.9931253 0.9532258 0.9540404 0.9708686
## [5,] 7.644348e+01 0.9931253 0.9715054 0.9585859 0.9772246
## [6,] 6.212030e+01 0.9920677 0.9768817 0.9590909 0.9788136
## [7,] 5.048084e+01 0.9910100 0.9790323 0.9631313 0.9777542
## [8,] 4.102226e+01 0.9899524 0.9795699 0.9636364 0.9782839
## [9,] 3.333594e+01 0.9899524 0.9806452 0.9641414 0.9804025
## [10,] 2.708979e+01 0.9910100 0.9860215 0.9681818 0.9830508
## [11,] 2.201399e+01 0.9915389 0.9876344 0.9702020 0.9819915
## [12,] 1.788924e+01 0.9920677 0.9897849 0.9742424 0.9814619
## [13,] 1.453734e+01 0.9931253 0.9919355 0.9757576 0.9804025
## [14,] 1.181348e+01 0.9952406 0.9930108 0.9767677 0.9804025
## [15,] 9.599994e+00 0.9957694 0.9935484 0.9813131 0.9798729
## [16,] 7.801246e+00 0.9962983 0.9940860 0.9833333 0.9793432
## [17,] 6.339530e+00 0.9968271 0.9940860 0.9858586 0.9793432
## [18,] 5.151694e+00 0.9968271 0.9946237 0.9873737 0.9793432
## [19,] 4.186423e+00 0.9968271 0.9946237 0.9883838 0.9798729

```

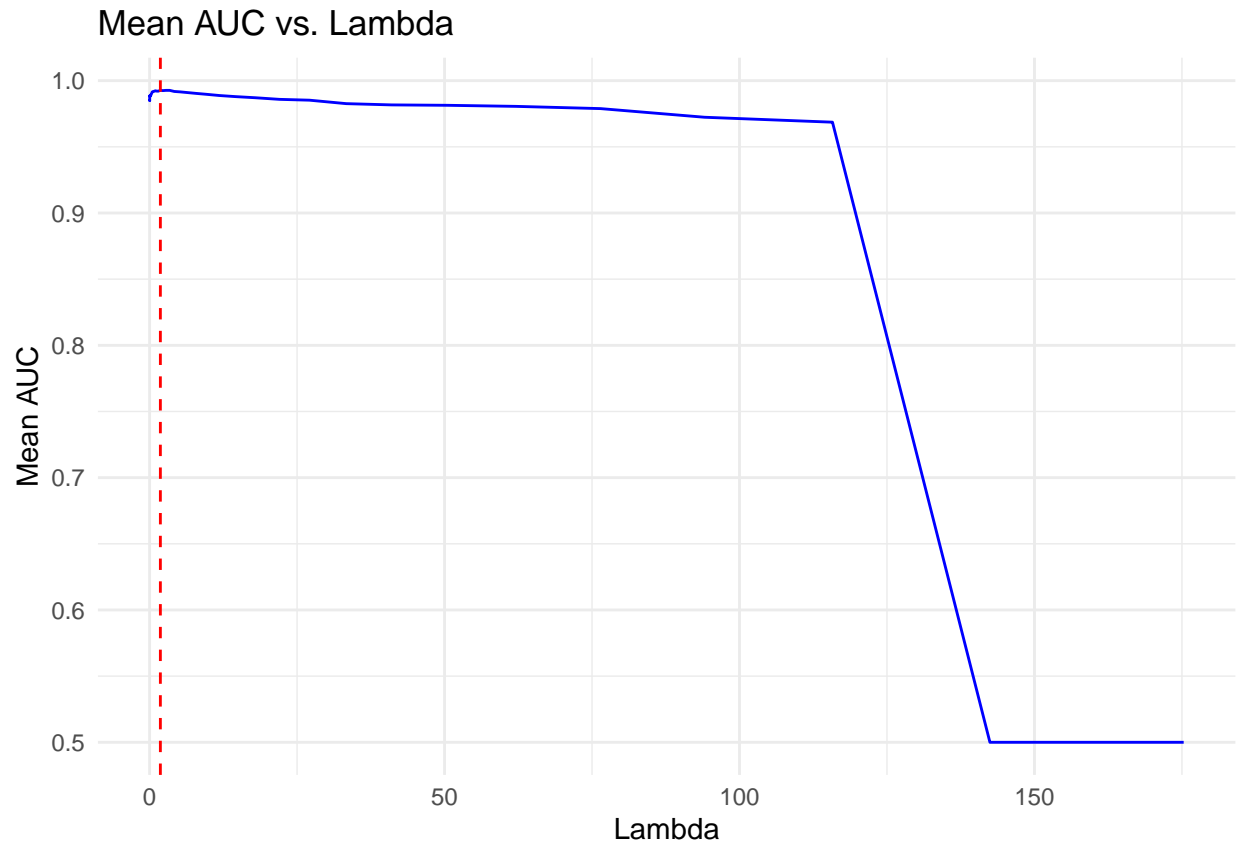
```

## [20,] 3.402015e+00 0.9968271 0.9962366 0.9898990 0.9804025
## [21,] 2.764580e+00 0.9968271 0.9962366 0.9909091 0.9804025
## [22,] 2.246582e+00 0.9968271 0.9962366 0.9904040 0.9804025
## [23,] 1.825641e+00 0.9962983 0.9962366 0.9898990 0.9825212
## [24,] 1.483571e+00 0.9962983 0.9962366 0.9868687 0.9830508
## [25,] 1.205595e+00 0.9957694 0.9962366 0.9858586 0.9846398
## [26,] 9.797031e-01 0.9957694 0.9978495 0.9848485 0.9856992
## [27,] 7.961365e-01 0.9957694 0.9978495 0.9843434 0.9867585
## [28,] 6.469647e-01 0.9952406 0.9973118 0.9833333 0.9878178
## [29,] 5.257431e-01 0.9957694 0.9978495 0.9823232 0.9878178
## [30,] 4.272348e-01 0.9952406 0.9973118 0.9813131 0.9883475
## [31,] 3.471840e-01 0.9931253 0.9967742 0.9787879 0.9888771
## [32,] 2.821323e-01 0.9925965 0.9967742 0.9772727 0.9894068
## [33,] 2.292692e-01 0.9925965 0.9951613 0.9757576 0.9894068
## [34,] 1.863111e-01 0.9931253 0.9946237 0.9737374 0.9904661
## [35,] 1.514021e-01 0.9931253 0.9935484 0.9712121 0.9909958
## [36,] 1.230339e-01 0.9936542 0.9935484 0.9696970 0.9904661
## [37,] 9.998112e-02 0.9931253 0.9935484 0.9681818 0.9909958
## [38,] 8.124769e-02 0.9920677 0.9935484 0.9666667 0.9899364
## [39,] 6.602434e-02 0.9920677 0.9919355 0.9676768 0.9899364
## [40,] 5.365338e-02 0.9915389 0.9924731 0.9676768 0.9920551
## [41,] 4.360037e-02 0.9894236 0.9919355 0.9671717 0.9931144
## [42,] 3.543098e-02 0.9894236 0.9919355 0.9671717 0.9931144
## [43,] 2.879229e-02 0.9894236 0.9919355 0.9671717 0.9936441
## [44,] 2.339749e-02 0.9894236 0.9913978 0.9661616 0.9931144
## [45,] 1.901351e-02 0.9746166 0.9913978 0.9661616 0.9931144
## [46,] 1.545096e-02 0.9746166 0.9913978 0.9661616 0.9936441
## [47,] 1.255592e-02 0.9746166 0.9913978 0.9656566 0.9936441
## [48,] 1.020332e-02 0.9746166 0.9913978 0.9656566 0.9936441
## [49,] 8.291527e-03 0.9746166 0.9919355 0.9661616 0.9936441
## [50,] 6.737947e-03 0.9746166 0.9919355 0.9661616 0.9936441
## [1] 0.5000000 0.5000000 0.9867367 0.9904905 0.9939940 0.9959960 0.9959960
## [8] 0.9969970 0.9979980 0.9979980 0.9979980 0.9979980 0.9984985 0.9984985
## [15] 0.9979980 0.9989990 0.9989990 0.9994995 0.9994995 1.0000000 0.9989990
## [22] 0.9984985 0.9984985 0.9979980 0.9979980 0.9974975 0.9954955 0.9949950
## [29] 0.9949950 0.9944945 0.9939940 0.9934935 0.9929930 0.9929930 0.9929930
## [36] 0.9974975 0.9984985 0.9994995 0.9994995 0.9994995 0.9994995 0.9994995
## [43] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [50] 1.0000000
##
##          res          auc          auc          auc          auc          auc
## [1,] 1.752945e+02 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000
## [2,] 1.424496e+02 0.5000000 0.5000000 0.5000000 0.5000000 0.5000000
## [3,] 1.157589e+02 0.9931253 0.9456989 0.9474747 0.9700742 0.9867367
## [4,] 9.406920e+01 0.9931253 0.9532258 0.9540404 0.9708686 0.9904905
## [5,] 7.644348e+01 0.9931253 0.9715054 0.9585859 0.9772246 0.9939940
## [6,] 6.212030e+01 0.9920677 0.9768817 0.9590909 0.9788136 0.9959960
## [7,] 5.048084e+01 0.9910100 0.9790323 0.9631313 0.9777542 0.9959960
## [8,] 4.102226e+01 0.9899524 0.9795699 0.9636364 0.9782839 0.9969970
## [9,] 3.333594e+01 0.9899524 0.9806452 0.9641414 0.9804025 0.9979980
## [10,] 2.708979e+01 0.9910100 0.9860215 0.9681818 0.9830508 0.9979980
## [11,] 2.201399e+01 0.9915389 0.9876344 0.9702020 0.9819915 0.9979980
## [12,] 1.788924e+01 0.9920677 0.9897849 0.9742424 0.9814619 0.9979980
## [13,] 1.453734e+01 0.9931253 0.9919355 0.9757576 0.9804025 0.9984985
## [14,] 1.181348e+01 0.9952406 0.9930108 0.9767677 0.9804025 0.9984985

```

```
## [15,] 9.599994e+00 0.9957694 0.9935484 0.9813131 0.9798729 0.9979980
## [16,] 7.801246e+00 0.9962983 0.9940860 0.9833333 0.9793432 0.9989990
## [17,] 6.339530e+00 0.9968271 0.9940860 0.9858586 0.9793432 0.9989990
## [18,] 5.151694e+00 0.9968271 0.9946237 0.9873737 0.9793432 0.9994995
## [19,] 4.186423e+00 0.9968271 0.9946237 0.9883838 0.9798729 0.9994995
## [20,] 3.402015e+00 0.9968271 0.9962366 0.9898990 0.9804025 1.0000000
## [21,] 2.764580e+00 0.9968271 0.9962366 0.9909091 0.9804025 0.9989990
## [22,] 2.246582e+00 0.9968271 0.9962366 0.9904040 0.9804025 0.9984985
## [23,] 1.825641e+00 0.9962983 0.9962366 0.9898990 0.9825212 0.9984985
## [24,] 1.483571e+00 0.9962983 0.9962366 0.9868687 0.9830508 0.9979980
## [25,] 1.205595e+00 0.9957694 0.9962366 0.9858586 0.9846398 0.9979980
## [26,] 9.797031e-01 0.9957694 0.9978495 0.9848485 0.9856992 0.9974975
## [27,] 7.961365e-01 0.9957694 0.9978495 0.9843434 0.9867585 0.9954955
## [28,] 6.469647e-01 0.9952406 0.9973118 0.9833333 0.9878178 0.9949950
## [29,] 5.257431e-01 0.9957694 0.9978495 0.9823232 0.9878178 0.9949950
## [30,] 4.272348e-01 0.9952406 0.9973118 0.9813131 0.9883475 0.9944945
## [31,] 3.471840e-01 0.9931253 0.9967742 0.9787879 0.9888771 0.9939940
## [32,] 2.821323e-01 0.9925965 0.9967742 0.9772727 0.9894068 0.9934935
## [33,] 2.292692e-01 0.9925965 0.9951613 0.9757576 0.9894068 0.9929930
## [34,] 1.863111e-01 0.9931253 0.9946237 0.9737374 0.9904661 0.9929930
## [35,] 1.514021e-01 0.9931253 0.9935484 0.9712121 0.9909958 0.9929930
## [36,] 1.230339e-01 0.9936542 0.9935484 0.9696970 0.9904661 0.9974975
## [37,] 9.998112e-02 0.9931253 0.9935484 0.9681818 0.9909958 0.9984985
## [38,] 8.124769e-02 0.9920677 0.9935484 0.9666667 0.9899364 0.9994995
## [39,] 6.602434e-02 0.9920677 0.9919355 0.9676768 0.9899364 0.9994995
## [40,] 5.365338e-02 0.9915389 0.9924731 0.9676768 0.9920551 0.9994995
## [41,] 4.360037e-02 0.9894236 0.9919355 0.9671717 0.9931144 0.9994995
## [42,] 3.543098e-02 0.9894236 0.9919355 0.9671717 0.9931144 0.9994995
## [43,] 2.879229e-02 0.9894236 0.9919355 0.9671717 0.9936441 1.0000000
## [44,] 2.339749e-02 0.9894236 0.9913978 0.9661616 0.9931144 1.0000000
## [45,] 1.901351e-02 0.9746166 0.9913978 0.9661616 0.9931144 1.0000000
## [46,] 1.545096e-02 0.9746166 0.9913978 0.9661616 0.9936441 1.0000000
## [47,] 1.255592e-02 0.9746166 0.9913978 0.9656566 0.9936441 1.0000000
## [48,] 1.020332e-02 0.9746166 0.9913978 0.9656566 0.9936441 1.0000000
## [49,] 8.291527e-03 0.9746166 0.9919355 0.9661616 0.9936441 1.0000000
## [50,] 6.737947e-03 0.9746166 0.9919355 0.9661616 0.9936441 1.0000000
```

```
cv_res <- as.data.frame(cv_test) #colnames(c("auc1", "auc2", "auc3", "auc4", "auc5"))
colnames(cv_res) <- c("res", "auc1", "auc2", "auc3", "auc4", "auc5")
cv_lambda <- cv_res[1]
mean_auc <- cv_res %>% dplyr::select(-1) %>% rowMeans()
cv_auc <- cbind(cv_lambda, mean_auc)
maxauc <- max(cv_auc$mean_auc)
bestlambda <- cv_auc[which(cv_auc$mean_auc == maxauc ),]$res
cv_auc %>%
  ggplot(x = res, y = mean_auc ) +
  geom_line(aes(x = res, y = mean_auc), col = "blue") +
  geom_vline(xintercept = bestlambda, linetype = "dashed", col = "red") +
  labs(title = "Mean AUC vs. Lambda",
       x = "Lambda",
       y = "Mean AUC")
```



## Compare full model and lasso model

```
#corresponding betas of best lambda
lasso_beta <- pathwise_sol[which(pathwise_sol$lambda == bestlambda ),][2:21] %>% as.numeric()

#prediction performance function
predict <- function(x, y, betavec) {
  theta <- x %*% betavec
  p <- exp(theta) / (1 + exp(theta))
  auc <- auc(y, p)
}

auc_lasso <- predict(x_test_stan, y_test, lasso_beta)
auc_lasso
```

```
## Area under the curve: 0.9958
```

```
cbind(auc_full, auc_lasso) %>% knitr::kable()
```

auc_full	auc_lasso
0.9771242	0.9957516



```
#coefficients of full and lasso models
glm_beta <- glm.fit$coefficients %>% as.vector()
coefnames <- rownames(coef(summary(glm.fit)))
cbind(coefnames, glm_beta, lasso_beta) %>% knitr::kable()
```

coefnames	glm_beta	lasso_beta
(Intercept)	-85.841618693509	-0.631891749918733
radius_mean	0.482809216338564	2.07950517324475
texture_mean	1.22782799498286	1.21495895057423
smoothness_mean	-61.2848250472609	0
concavity_mean	277.147953900727	0.915894304171588
symmetry_mean	-174.873046711938	0
fractal_dimension_mean	-268.503538734735	0
radius_se	68.8555271776555	2.23661374208485
texture_se	-0.498284713431817	0
smoothness_se	1374.2230814889	0
compactness_se	42.7436942672398	-0.527415419852414
concavity_se	-150.677448596641	0
concave_points_se	731.297840302427	0
symmetry_se	-1308.68956822617	-0.385186799354295
fractal_dimension_se	-6434.27652769617	-0.421464230008035
smoothness_worst	-133.987895735445	0.464232406689049
compactness_worst	-104.622145944305	0
concave_points_worst	161.993778687915	2.1144068366775
symmetry_worst	180.871288833035	0.745540291717346
fractal_dimension_worst	807.843869038758	0