**MIDTERM: MAT 443**

**Name: ERIC AGYEMANG**

**Question 1.**

**(a)**

**The output below is the full model including all necessary variables with both significant variables and non-significant variables.**

> lm.fit1 = lm(nature~.,data=jmtrain)

> summary(lm.fit1)

Call:

lm(formula = nature ~ ., data = jmtrain)

Residuals:

Min 1Q Median 3Q Max

-1.33886 -0.03718 -0.00904 0.01823 0.97289

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.528e-02 4.184e-02 -0.604 0.54572

x1 1.690e-03 1.352e-03 1.250 0.21138

x2 -1.427e-03 1.333e-03 -1.071 0.28434

x3 4.567e-03 4.112e-03 1.111 0.26680

x4 3.099e-03 1.724e-03 1.797 0.07236 .

x5 1.694e-07 1.014e-07 1.670 0.09490 .

x6 -6.135e-08 1.878e-07 -0.327 0.74389

x7 -1.576e-07 1.076e-07 -1.464 0.14324

x8 -5.654e-06 5.556e-07 -10.177 < 2e-16 \*\*\*

x9 -4.455e-06 1.251e-05 -0.356 0.72167

x10 2.106e-05 1.583e-05 1.330 0.18343

x11 1.741e-05 2.246e-05 0.775 0.43834

x12 -2.570e-05 1.306e-05 -1.967 0.04923 \*

x13 6.466e-05 2.296e-03 0.028 0.97753

x14 -8.642e-03 4.057e-03 -2.130 0.03318 \*

x15 -8.460e-07 3.672e-07 -2.304 0.02123 \*

x16 2.034e-08 1.470e-08 1.384 0.16644

x17 5.886e-05 5.031e-05 1.170 0.24208

x18 1.048e-04 3.566e-05 2.939 0.00330 \*\*

x19 -1.321e-04 4.753e-05 -2.778 0.00547 \*\*

x42 -4.801e-04 4.128e-03 -0.116 0.90741

x43 1.715e-04 2.485e-03 0.069 0.94496

x44 3.932e-03 1.410e-03 2.789 0.00529 \*\*

x45 -5.653e-04 5.130e-04 -1.102 0.27057

x46 -1.794e-03 4.065e-03 -0.441 0.65892

x47 -3.851e-08 2.176e-08 -1.770 0.07683 .

x48 -1.150e-03 4.095e-03 -0.281 0.77889

x49 3.311e-04 4.134e-03 0.080 0.93617

x50 6.725e-03 4.108e-03 1.637 0.10164

x51 -1.085e-03 4.049e-03 -0.268 0.78873

x52 1.670e-03 4.096e-03 0.408 0.68350

x53 -6.724e-03 4.090e-03 -1.644 0.10019

x54 -3.656e-03 4.083e-03 -0.896 0.37052

x55 2.928e-03 4.117e-03 0.711 0.47696

x56 -1.085e-02 2.347e-03 -4.623 3.83e-06 \*\*\*

x57 3.930e-03 4.114e-03 0.955 0.33948

x58 -3.066e-03 4.013e-03 -0.764 0.44488

x59 -1.050e-03 4.048e-03 -0.259 0.79531

x60 2.954e-03 4.070e-03 0.726 0.46798

x61 2.812e-04 4.153e-03 0.068 0.94601

x62 1.596e-03 1.408e-03 1.134 0.25693

x63 1.758e-02 2.121e-03 8.291 < 2e-16 \*\*\*

x64 4.584e-03 2.252e-03 2.035 0.04186 \*

x65 1.487e-03 4.035e-03 0.368 0.71258

x66 -8.409e-03 4.032e-03 -2.086 0.03702 \*

x67 2.861e-03 3.547e-03 0.807 0.41989

x68 2.267e-03 4.144e-03 0.547 0.58437

x69 8.021e-03 1.377e-03 5.826 5.84e-09 \*\*\*

x70 -1.036e-04 4.150e-04 -0.250 0.80295

x71 2.045e-05 4.731e-05 0.432 0.66558

x72 1.534e-04 4.421e-04 0.347 0.72857

x73 -5.134e-03 2.724e-03 -1.885 0.05950 .

x74 -1.781e-03 4.178e-03 -0.426 0.66983

x75 -6.850e-04 6.893e-04 -0.994 0.32034

x76 -3.183e-04 7.256e-04 -0.439 0.66092

x77 -2.971e-03 2.274e-03 -1.307 0.19134

x78 -2.321e-03 8.062e-04 -2.878 0.00401 \*\*

x79 2.960e-04 3.620e-03 0.082 0.93483

x80 -1.233e-03 1.965e-03 -0.628 0.53031

x81 -1.191e-03 1.869e-03 -0.637 0.52388

x82 3.919e-04 2.750e-03 0.143 0.88666

x83 -2.144e-03 1.135e-03 -1.889 0.05888 .

x84 2.674e-04 8.093e-04 0.330 0.74112

x85 2.095e-05 2.486e-04 0.084 0.93284

x86 -1.213e-05 8.154e-05 -0.149 0.88170

x87 -5.601e-04 4.705e-04 -1.190 0.23391

x88 1.130e-04 6.727e-05 1.679 0.09316 .

x89 -2.714e-03 2.231e-03 -1.217 0.22381

x90 6.898e-04 9.142e-04 0.755 0.45056

x91 -1.884e-03 1.690e-03 -1.115 0.26492

x92 -3.550e-04 1.026e-03 -0.346 0.72930

x93 -1.044e-04 2.813e-03 -0.037 0.97040

x94 -1.906e-05 5.428e-05 -0.351 0.72553

x95 3.062e-03 3.676e-03 0.833 0.40489

x96 -5.306e-05 2.437e-04 -0.218 0.82768

x97 -2.542e-05 8.486e-04 -0.030 0.97610

x98 -1.729e-03 2.050e-03 -0.843 0.39911

x99 -8.508e-04 3.110e-03 -0.274 0.78445

x100 1.836e-03 3.936e-03 0.466 0.64096

x101 1.155e-05 9.014e-05 0.128 0.89801

x102 -8.925e-05 2.233e-04 -0.400 0.68943

x103 -1.797e-05 8.782e-05 -0.205 0.83791

x104 -1.164e-04 1.905e-04 -0.611 0.54108

x105 -6.547e-04 2.968e-03 -0.221 0.82544

x106 -8.589e-04 1.490e-03 -0.576 0.56439

x107 -1.166e-04 2.816e-04 -0.414 0.67884

x108 3.431e-03 3.898e-03 0.880 0.37880

x109 -2.085e-04 1.555e-04 -1.341 0.18010

x110 1.423e-03 1.449e-03 0.982 0.32590

x111 -5.369e-04 4.478e-04 -1.199 0.23052

x112 -4.489e-04 5.040e-04 -0.891 0.37312

x113 -2.571e-05 1.040e-04 -0.247 0.80481

x114 -2.574e-04 7.011e-04 -0.367 0.71357

x115 1.056e-04 1.878e-04 0.563 0.57377

x116 -4.132e-05 5.034e-04 -0.082 0.93459

x117 3.178e-04 8.753e-04 0.363 0.71653

x118 3.313e-05 1.075e-04 0.308 0.75793

x119 2.613e-05 2.351e-05 1.111 0.26650

x120 -6.025e-05 1.570e-04 -0.384 0.70123

x121 -1.761e-05 1.544e-05 -1.140 0.25414

x122 -3.258e-05 5.746e-05 -0.567 0.57066

x123 1.055e-04 6.769e-05 1.559 0.11898

x124 -6.521e-03 3.995e-03 -1.632 0.10267

x201 4.446e-03 1.405e-03 3.165 0.00155 \*\*

x202 -2.868e-03 1.246e-03 -2.302 0.02138 \*

x203 2.258e-01 4.251e-03 53.126 < 2e-16 \*\*\*

x204 3.011e-03 1.654e-03 1.820 0.06881 .

x205 4.991e-03 1.668e-03 2.993 0.00277 \*\*

x206 -1.292e-02 3.935e-03 -3.284 0.00103 \*\*

x207 2.341e-01 5.333e-03 43.900 < 2e-16 \*\*\*

x208 6.682e-03 1.389e-03 4.810 1.53e-06 \*\*\*

x209 2.133e-01 7.053e-03 30.242 < 2e-16 \*\*\*

x210 -5.621e-03 1.265e-03 -4.442 9.01e-06 \*\*\*

x211 1.442e-02 2.554e-03 5.648 1.67e-08 \*\*\*

x212 7.089e-03 1.630e-03 4.348 1.39e-05 \*\*\*

x213 4.233e-02 3.327e-03 12.726 < 2e-16 \*\*\*

x214 2.037e-02 3.584e-03 5.685 1.35e-08 \*\*\*

x215 -1.235e-03 1.672e-03 -0.739 0.46013

x216 -2.424e-03 1.317e-03 -1.840 0.06577 .

x217 -5.173e-04 1.328e-03 -0.390 0.69680

x218 1.986e-01 5.734e-03 34.636 < 2e-16 \*\*\*

x219 1.265e-03 2.734e-04 4.626 3.78e-06 \*\*\*

x220 -1.643e-02 4.168e-03 -3.941 8.15e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1173 on 9937 degrees of freedom

Multiple R-squared: 0.7409, Adjusted R-squared: 0.7378

F-statistic: 233 on 122 and 9937 DF, p-value: < 2.2e-16

> anova(lm.fit1)

Analysis of Variance Table

Response: nature

Df Sum Sq Mean Sq F value Pr(>F)

x1 1 3.726 3.726 270.6918 < 2.2e-16 \*\*\*

x2 1 0.011 0.011 0.7806 0.3769817

x3 1 0.016 0.016 1.1950 0.2743485

x4 1 0.000 0.000 0.0136 0.9070073

x5 1 0.302 0.302 21.9623 2.818e-06 \*\*\*

x6 1 2.092 2.092 151.9749 < 2.2e-16 \*\*\*

x7 1 0.028 0.028 2.0663 0.1506167

x8 1 51.974 51.974 3776.2680 < 2.2e-16 \*\*\*

x9 1 0.372 0.372 26.9939 2.081e-07 \*\*\*

x10 1 0.012 0.012 0.8817 0.3477706

x11 1 0.012 0.012 0.9049 0.3414986

x12 1 0.055 0.055 3.9664 0.0464454 \*

x13 1 0.012 0.012 0.8759 0.3493407

x14 1 0.106 0.106 7.7225 0.0054640 \*\*

x15 1 0.088 0.088 6.3813 0.0115481 \*

x16 1 0.000 0.000 0.0002 0.9901927

x17 1 0.006 0.006 0.4540 0.5004365

x18 1 0.325 0.325 23.6313 1.185e-06 \*\*\*

x19 1 0.157 0.157 11.4227 0.0007283 \*\*\*

x42 1 0.003 0.003 0.2238 0.6361572

x43 1 0.033 0.033 2.4125 0.1204033

x44 1 40.912 40.912 2972.5143 < 2.2e-16 \*\*\*

x45 1 0.023 0.023 1.6457 0.1995750

x46 1 0.099 0.099 7.1746 0.0074064 \*\*

x47 1 2.219 2.219 161.2358 < 2.2e-16 \*\*\*

x48 1 0.137 0.137 9.9501 0.0016132 \*\*

x49 1 0.003 0.003 0.2115 0.6455709

x50 1 0.022 0.022 1.5621 0.2113811

x51 1 0.013 0.013 0.9333 0.3340384

x52 1 0.008 0.008 0.5677 0.4511983

x53 1 0.001 0.001 0.0399 0.8416339

x54 1 0.076 0.076 5.4893 0.0191525 \*

x55 1 0.031 0.031 2.2614 0.1326674

x56 1 1.933 1.933 140.4201 < 2.2e-16 \*\*\*

x57 1 0.003 0.003 0.2515 0.6160040

x58 1 0.000 0.000 0.0327 0.8564752

x59 1 0.000 0.000 0.0047 0.9452888

x60 1 0.001 0.001 0.0898 0.7644129

x61 1 0.001 0.001 0.0906 0.7633955

x62 1 0.546 0.546 39.7025 3.083e-10 \*\*\*

x63 1 11.656 11.656 846.9112 < 2.2e-16 \*\*\*

x64 1 0.364 0.364 26.4527 2.752e-07 \*\*\*

x65 1 0.058 0.058 4.1966 0.0405307 \*

x66 1 0.034 0.034 2.4606 0.1167689

x67 1 0.006 0.006 0.4151 0.5194105

x68 1 0.083 0.083 6.0414 0.0139911 \*

x69 1 2.522 2.522 183.2426 < 2.2e-16 \*\*\*

x70 1 0.040 0.040 2.9375 0.0865759 .

x71 1 0.026 0.026 1.8848 0.1698249

x72 1 0.008 0.008 0.5924 0.4415278

x73 1 0.105 0.105 7.6506 0.0056857 \*\*

x74 1 0.064 0.064 4.6190 0.0316430 \*

x75 1 0.157 0.157 11.3961 0.0007388 \*\*\*

x76 1 0.064 0.064 4.6304 0.0314336 \*

x77 1 0.054 0.054 3.9322 0.0473973 \*

x78 1 0.419 0.419 30.4300 3.548e-08 \*\*\*

x79 1 0.000 0.000 0.0124 0.9114373

x80 1 0.152 0.152 11.0418 0.0008940 \*\*\*

x81 1 0.060 0.060 4.3518 0.0369958 \*

x82 1 0.009 0.009 0.6384 0.4242936

x83 1 0.260 0.260 18.8671 1.415e-05 \*\*\*

x84 1 0.002 0.002 0.1502 0.6983169

x85 1 0.000 0.000 0.0258 0.8724690

x86 1 0.011 0.011 0.8241 0.3640047

x87 1 0.088 0.088 6.4223 0.0112851 \*

x88 1 0.384 0.384 27.9193 1.292e-07 \*\*\*

x89 1 0.011 0.011 0.7906 0.3739331

x90 1 0.000 0.000 0.0222 0.8815208

x91 1 0.092 0.092 6.6866 0.0097280 \*\*

x92 1 0.102 0.102 7.4362 0.0064038 \*\*

x93 1 0.002 0.002 0.1363 0.7120300

x94 1 0.001 0.001 0.0383 0.8447790

x95 1 0.002 0.002 0.1332 0.7151892

x96 1 0.007 0.007 0.5134 0.4736998

x97 1 0.004 0.004 0.2984 0.5848681

x98 1 0.039 0.039 2.8107 0.0936663 .

x99 1 0.001 0.001 0.0867 0.7684684

x100 1 0.003 0.003 0.2017 0.6533940

x101 1 0.016 0.016 1.1405 0.2855798

x102 1 0.005 0.005 0.3377 0.5611759

x103 1 0.081 0.081 5.8490 0.0156035 \*

x104 1 0.082 0.082 5.9626 0.0146299 \*

x105 1 0.093 0.093 6.7716 0.0092755 \*\*

x106 1 0.008 0.008 0.5504 0.4581828

x107 1 0.002 0.002 0.1185 0.7306761

x108 1 0.023 0.023 1.6980 0.1925832

x109 1 0.494 0.494 35.9231 2.124e-09 \*\*\*

x110 1 0.204 0.204 14.8013 0.0001202 \*\*\*

x111 1 0.329 0.329 23.8973 1.032e-06 \*\*\*

x112 1 0.089 0.089 6.4557 0.0110748 \*

x113 1 0.041 0.041 2.9811 0.0842719 .

x114 1 0.035 0.035 2.5317 0.1116136

x115 1 0.000 0.000 0.0005 0.9819558

x116 1 0.024 0.024 1.7696 0.1834650

x117 1 0.147 0.147 10.6973 0.0010766 \*\*

x118 1 0.007 0.007 0.5058 0.4769813

x119 1 0.149 0.149 10.8579 0.0009872 \*\*\*

x120 1 0.041 0.041 2.9707 0.0848158 .

x121 1 0.068 0.068 4.9626 0.0259238 \*

x122 1 0.064 0.064 4.6694 0.0307288 \*

x123 1 0.022 0.022 1.5718 0.2099796

x124 1 0.011 0.011 0.7863 0.3752523

x201 1 1.184 1.184 86.0349 < 2.2e-16 \*\*\*

x202 1 1.128 1.128 81.9740 < 2.2e-16 \*\*\*

x203 1 166.952 166.952 12130.2437 < 2.2e-16 \*\*\*

x204 1 0.051 0.051 3.6758 0.0552375 .

x205 1 0.274 0.274 19.9035 8.235e-06 \*\*\*

x206 1 0.494 0.494 35.8643 2.189e-09 \*\*\*

x207 1 55.248 55.248 4014.1112 < 2.2e-16 \*\*\*

x208 1 0.536 0.536 38.9779 4.462e-10 \*\*\*

x209 1 18.928 18.928 1375.2206 < 2.2e-16 \*\*\*

x210 1 0.412 0.412 29.9252 4.599e-08 \*\*\*

x211 1 0.563 0.563 40.9122 1.664e-10 \*\*\*

x212 1 0.402 0.402 29.2155 6.626e-08 \*\*\*

x213 1 3.014 3.014 218.9576 < 2.2e-16 \*\*\*

x214 1 0.386 0.386 28.0247 1.223e-07 \*\*\*

x215 1 0.020 0.020 1.4522 0.2282081

x216 1 0.051 0.051 3.7062 0.0542398 .

x217 1 0.019 0.019 1.3984 0.2370178

x218 1 16.719 16.719 1214.7387 < 2.2e-16 \*\*\*

x219 1 0.295 0.295 21.4593 3.660e-06 \*\*\*

x220 1 0.214 0.214 15.5350 8.155e-05 \*\*\*

Residuals 9937 136.766 0.014

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**(b)**

**There were 29 out 122 predictors significant variables obtain from the results. This is called the reduced model, which contain the dropped non-significant terms. Even though these two model that is, full model and reduce model are relatively statistically equivalent through the format test developed below. We use a significance level of 0.05 to elaborate between significant and non-significant variables.**

> lm.fit2 = lm(nature ~ x8+x12+x14+x15+x18+x19+x44+x56+x63+x64+x66+x69+x78+x201+x202+x203+x205+x206+x207+x208+x209+x210+x211+x212+x213+x214+x218+x219+x220,data = jmtrain)

> summary(lm.fit2)

Call:

lm(formula = nature ~ x8 + x12 + x14 + x15 + x18 + x19 + x44 +

x56 + x63 + x64 + x66 + x69 + x78 + x201 + x202 + x203 +

x205 + x206 + x207 + x208 + x209 + x210 + x211 + x212 + x213 +

x214 + x218 + x219 + x220, data = jmtrain)

Residuals:

Min 1Q Median 3Q Max

-1.34434 -0.03690 -0.00892 0.01750 0.95492

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -7.169e-02 7.908e-03 -9.065 < 2e-16 \*\*\*

x8 -5.524e-06 5.204e-07 -10.614 < 2e-16 \*\*\*

x12 6.602e-07 4.066e-07 1.624 0.104458

x14 -8.423e-03 4.037e-03 -2.087 0.036945 \*

x15 -8.837e-07 3.639e-07 -2.428 0.015184 \*

x18 9.910e-05 3.447e-05 2.875 0.004052 \*\*

x19 -1.241e-04 4.575e-05 -2.711 0.006711 \*\*

x44 5.527e-03 2.847e-04 19.411 < 2e-16 \*\*\*

x56 -1.113e-02 2.333e-03 -4.772 1.85e-06 \*\*\*

x63 1.772e-02 2.108e-03 8.406 < 2e-16 \*\*\*

x64 4.995e-03 2.241e-03 2.229 0.025864 \*

x66 -7.549e-03 4.014e-03 -1.881 0.060042 .

x69 8.019e-03 1.371e-03 5.851 5.04e-09 \*\*\*

x78 -2.720e-03 7.959e-04 -3.417 0.000635 \*\*\*

x201 4.344e-03 1.398e-03 3.108 0.001892 \*\*

x202 -2.932e-03 1.241e-03 -2.363 0.018142 \*

x203 2.276e-01 4.237e-03 53.724 < 2e-16 \*\*\*

x205 4.933e-03 1.659e-03 2.974 0.002946 \*\*

x206 -1.313e-02 3.909e-03 -3.358 0.000787 \*\*\*

x207 2.348e-01 5.321e-03 44.124 < 2e-16 \*\*\*

x208 6.627e-03 1.382e-03 4.796 1.64e-06 \*\*\*

x209 2.141e-01 7.029e-03 30.461 < 2e-16 \*\*\*

x210 -5.766e-03 1.258e-03 -4.582 4.67e-06 \*\*\*

x211 1.416e-02 2.540e-03 5.575 2.53e-08 \*\*\*

x212 7.070e-03 1.624e-03 4.354 1.35e-05 \*\*\*

x213 4.278e-02 3.313e-03 12.915 < 2e-16 \*\*\*

x214 2.078e-02 3.569e-03 5.822 5.98e-09 \*\*\*

x218 1.994e-01 5.719e-03 34.865 < 2e-16 \*\*\*

x219 1.295e-03 2.723e-04 4.756 2.00e-06 \*\*\*

x220 -1.654e-02 4.150e-03 -3.986 6.76e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1173 on 10030 degrees of freedom

Multiple R-squared: 0.7384, Adjusted R-squared: 0.7377

F-statistic: 976.5 on 29 and 10030 DF, p-value: < 2.2e-16

> anova(lm.fit1,lm.fit2)

Analysis of Variance Table

Model 1: nature ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 +

x12 + x13 + x14 + x15 + x16 + x17 + x18 + x19 + x42 + x43 +

x44 + x45 + x46 + x47 + x48 + x49 + x50 + x51 + x52 + x53 +

x54 + x55 + x56 + x57 + x58 + x59 + x60 + x61 + x62 + x63 +

x64 + x65 + x66 + x67 + x68 + x69 + x70 + x71 + x72 + x73 +

x74 + x75 + x76 + x77 + x78 + x79 + x80 + x81 + x82 + x83 +

x84 + x85 + x86 + x87 + x88 + x89 + x90 + x91 + x92 + x93 +

x94 + x95 + x96 + x97 + x98 + x99 + x100 + x101 + x102 +

x103 + x104 + x105 + x106 + x107 + x108 + x109 + x110 + x111 +

x112 + x113 + x114 + x115 + x116 + x117 + x118 + x119 + x120 +

x121 + x122 + x123 + x124 + x201 + x202 + x203 + x204 + x205 +

x206 + x207 + x208 + x209 + x210 + x211 + x212 + x213 + x214 +

x215 + x216 + x217 + x218 + x219 + x220

Model 2: nature ~ x8 + x12 + x14 + x15 + x18 + x19 + x44 + x56 + x63 +

x64 + x66 + x69 + x78 + x201 + x202 + x203 + x205 + x206 +

x207 + x208 + x209 + x210 + x211 + x212 + x213 + x214 + x218 +

x219 + x220

Res.Df RSS Df Sum of Sq F Pr(>F)

1 9937 136.77

2 10030 138.08 -93 -1.3168 1.0288 0.4046

**(c)**

**CONFUSION MATRIX (MULTIPLE REGRESSION)**

> probabilities = predict(lm.fit2, newdata = jmtest)

> pred = rep("0", nrow(jmtest))

> pred[probabilities > 0.5] = "1"

> table(pred, jmtest$nature)

pred 0 1

0 7485 134

1 16 433

**CALCULATIING OF FP, FN, SENSITIVITY AND SPECIFICITY RATES (MULTIPLE REGRESSION)**

**FP rate =**

**FN rate=**

**SENSITIVITY =**

**SPECIFICITY=**

**MISCLASSIFICATION RATE = 0.0185%**

**Looking at above rates, sensitivity records the highest rest in this multiple regression model which is 98.41 percent, hence there is a malicious detection on the website is accurate because of the conclusion above.**

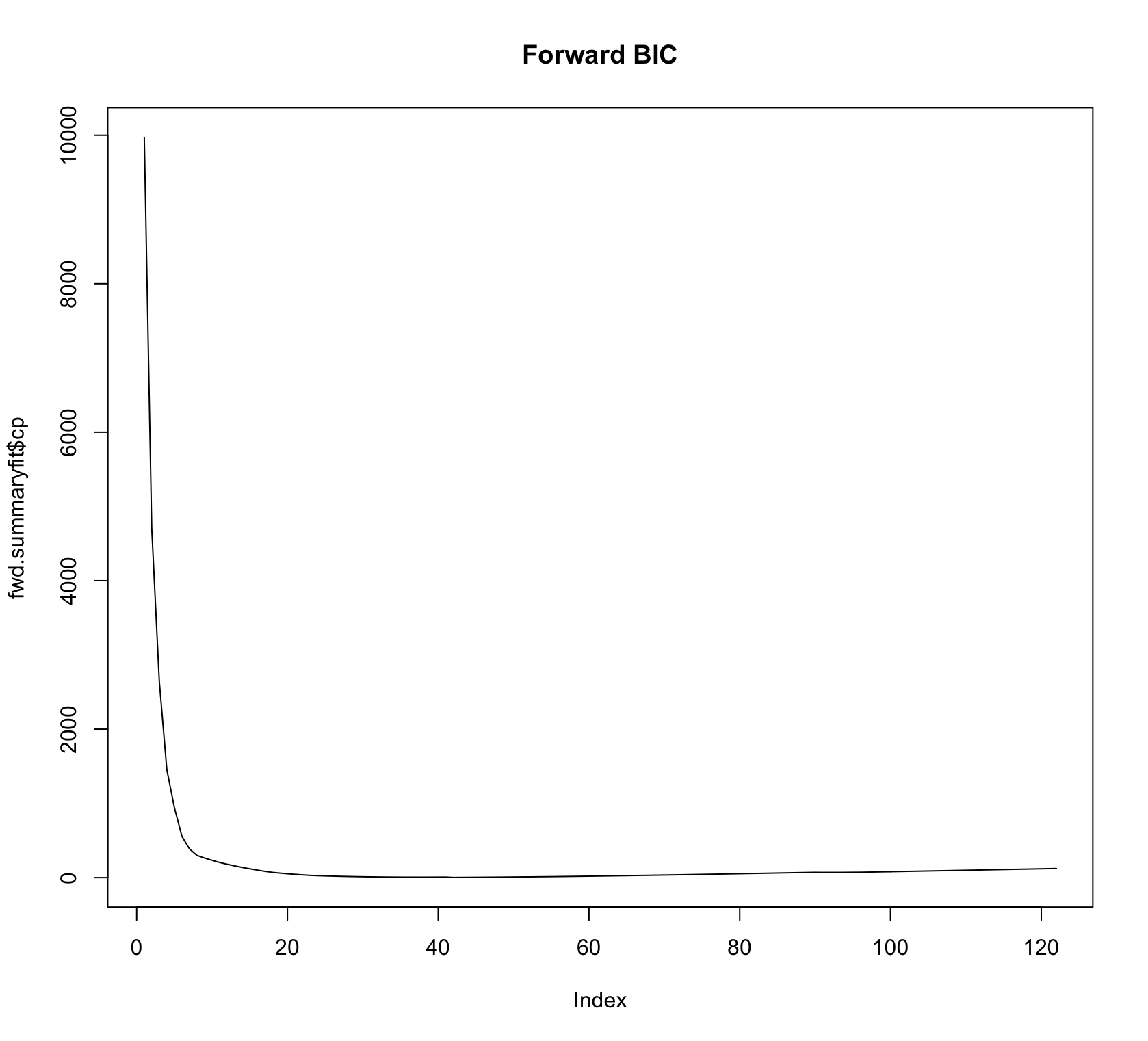
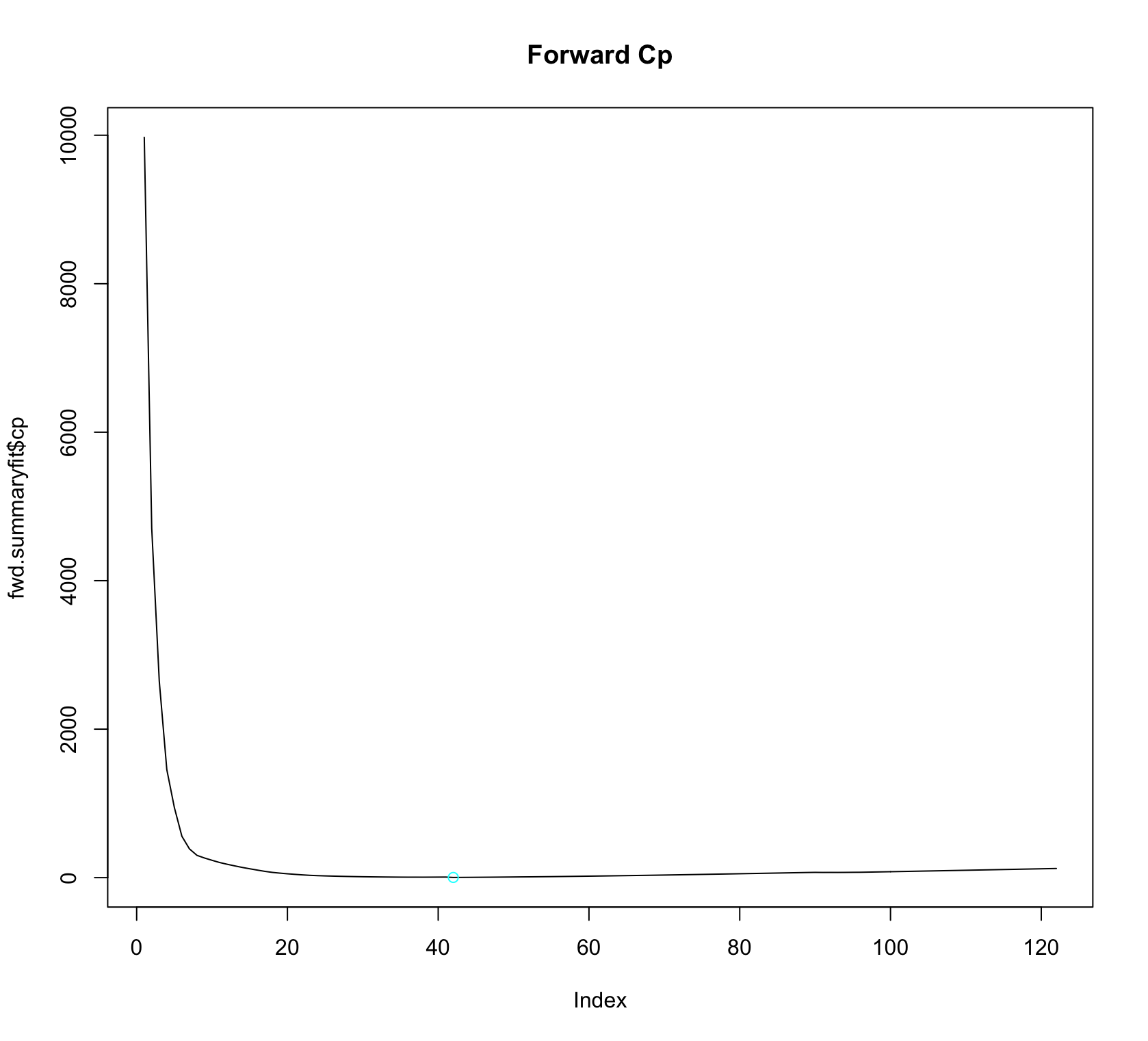
**Question 2.**

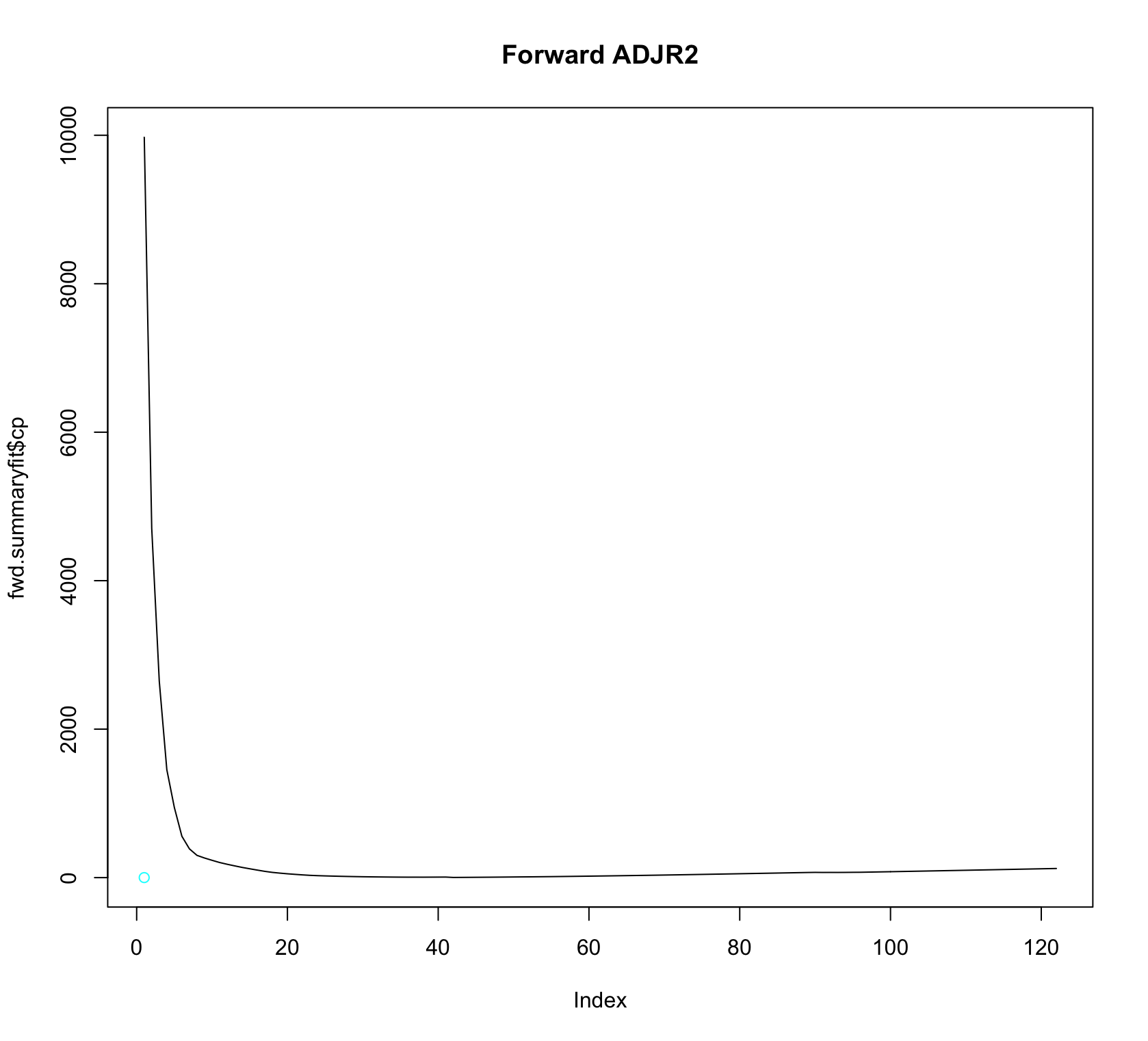
**(a)**

**i. Forward Methods**

> fit.fwd = regsubsets(nature~.,data=jmtrain,nvmax=122,method="forward")

> fwd.summaryfit = summary(fit.fwd)

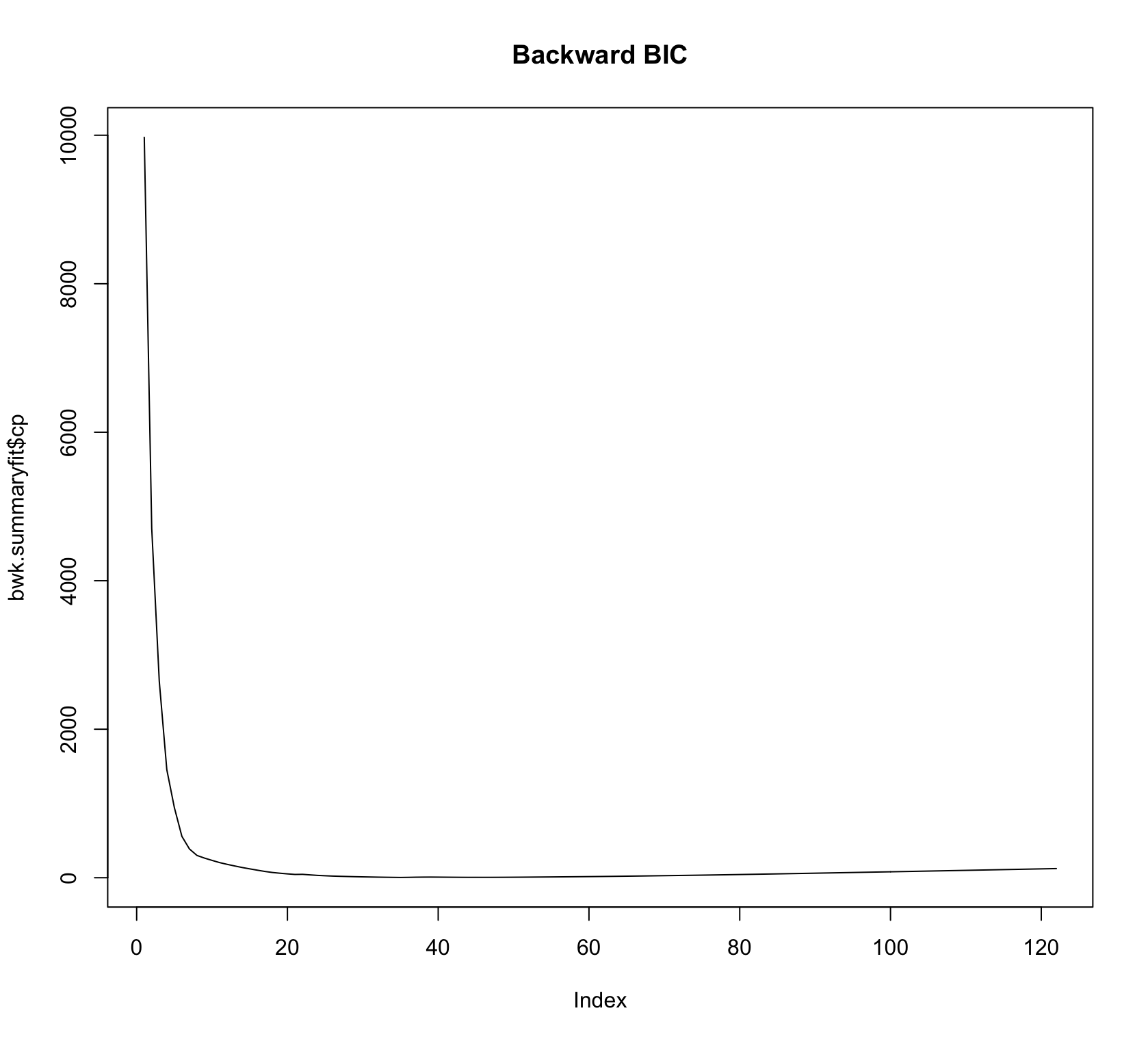
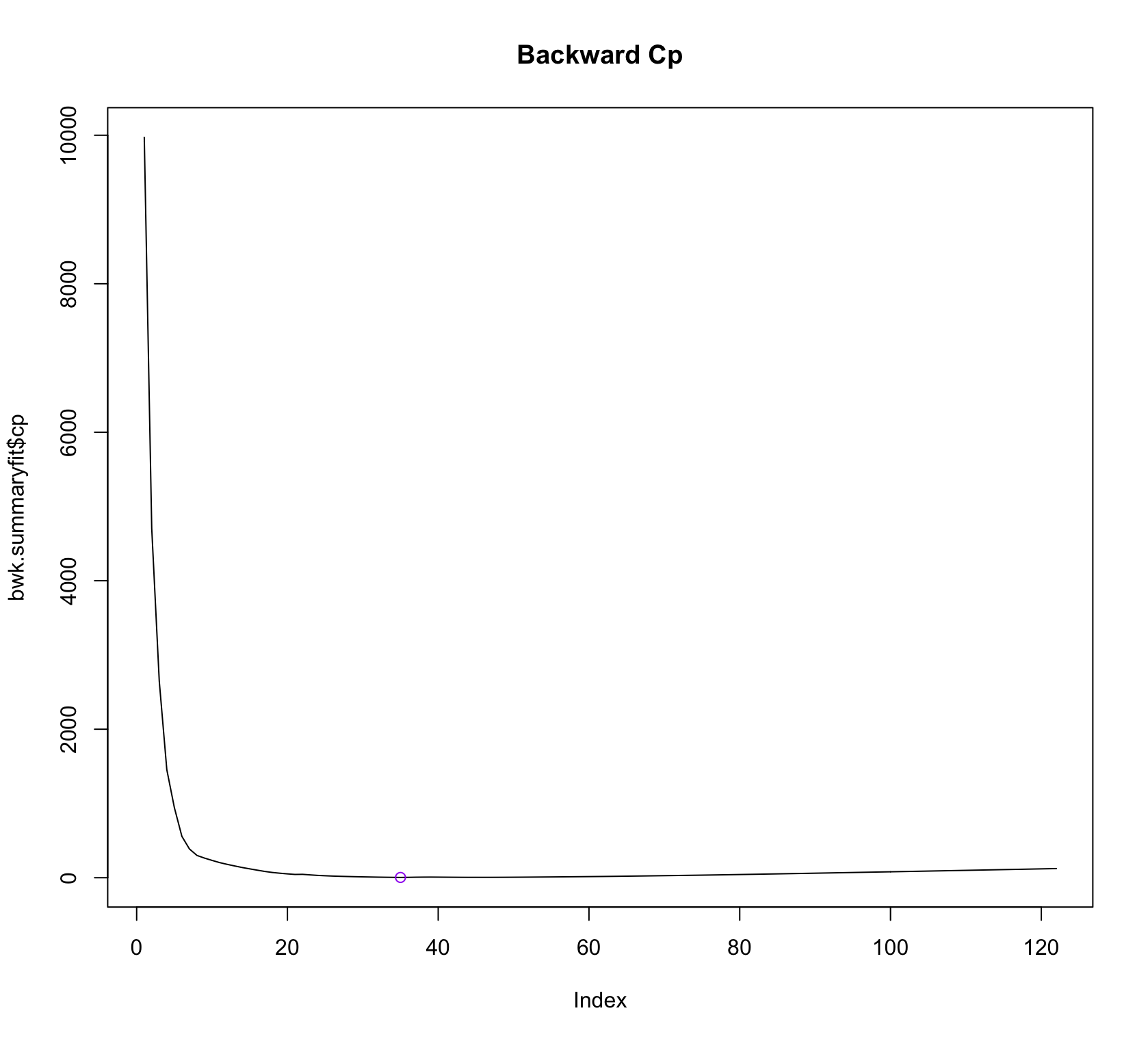


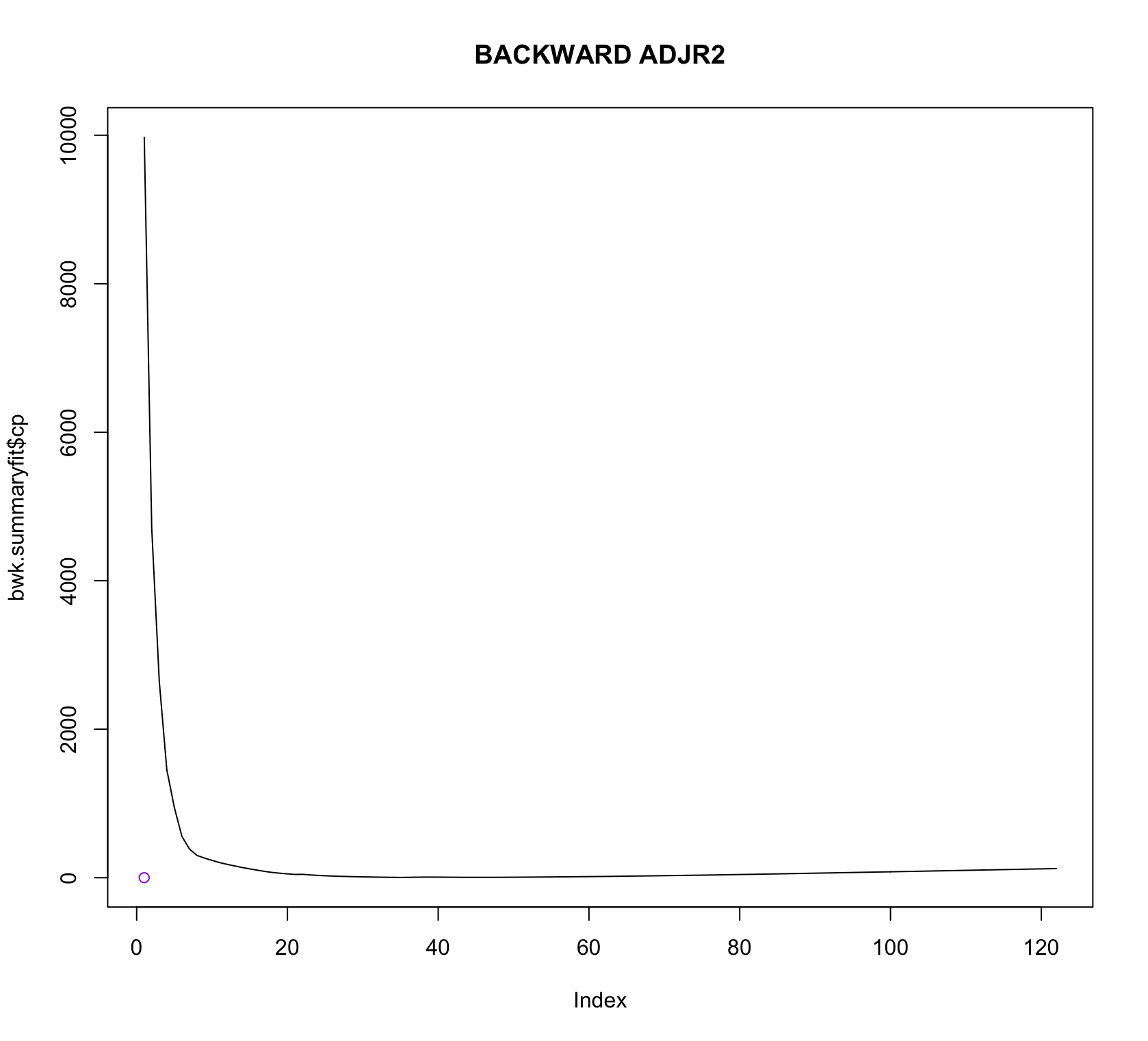


**ii. Backward Methods**

> fit.bwk = regsubsets(nature~.,data=jmtrain,nvmax=122,method="backward")

> bwk.summaryfit=summary(fit.bwk)





iii. Forward stepwise selection

> which.min(fwd.summaryfit$cp)

[1] 42

> which.min(fwd.summaryfit$bic)

[1] 21

> which.min(fwd.summaryfit$adjr2)

[1] 1

iv.Backward stepwise selection

> which.min(bwk.summaryfit$cp)

[1] 35

> which.min(bwk.summaryfit$bic)

[1] 21

> which.min(bwk.summaryfit$adjr2)

[1] 1

**Since forward and backward stepwise produce different measures of best model selection, BIC and adjusted R square are shown for the best models of each above. The training data set (the lower frontier).Cp and BIC are estimate of test MSE. In the middle of plot we see the the BIC estimate of test error shows an increase after the variables selected. The other two plots are rather flat after some variables are added.**

**(b)**

**i.**

we shall now get our coefficient for 21 variables model for both stepwise selection

> coef(fit.fwd,21)

(Intercept) x8 x44 x47

-6.095649e-02 -6.145621e-06 5.460678e-03 -6.199597e-08

x56 x63 x69 x78

-1.084668e-02 1.797753e-02 8.363666e-03 -2.638221e-03

x203 x205 x206 x207

2.275970e-01 5.139366e-03 -1.324550e-02 2.347763e-01

x208 x209 x210 x211

6.765808e-03 2.143590e-01 -5.653449e-03 1.398682e-02

x212 x213 x214 x218

7.125842e-03 4.297744e-02 2.021660e-02 1.994953e-01

x219 x220

1.297539e-03 -1.654728e-02

> lm.fit3 = lm(nature~x8+x44+x47+x56+x63+x69+x78+x203+x205+x206+x207+x208+x209

+ +x210+x211+x212+x213+x214+x218+x219+x220, data=jmtrain)

> summary(lm.fit3)

Call:

lm(formula = nature ~ x8 + x44 + x47 + x56 + x63 + x69 + x78 +

x203 + x205 + x206 + x207 + x208 + x209 + x210 + x211 + x212 +

x213 + x214 + x218 + x219 + x220, data = jmtrain)

Residuals:

Min 1Q Median 3Q Max

-1.32843 -0.03658 -0.00875 0.01740 0.97694

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -6.096e-02 7.335e-03 -8.311 < 2e-16 \*\*\*

x8 -6.146e-06 3.308e-07 -18.576 < 2e-16 \*\*\*

x44 5.461e-03 2.836e-04 19.252 < 2e-16 \*\*\*

x47 -6.200e-08 1.464e-08 -4.234 2.32e-05 \*\*\*

x56 -1.085e-02 2.333e-03 -4.649 3.38e-06 \*\*\*

x63 1.798e-02 2.109e-03 8.522 < 2e-16 \*\*\*

x69 8.364e-03 1.364e-03 6.131 9.09e-10 \*\*\*

x78 -2.638e-03 7.952e-04 -3.318 0.000911 \*\*\*

x203 2.276e-01 4.239e-03 53.697 < 2e-16 \*\*\*

x205 5.139e-03 1.659e-03 3.097 0.001960 \*\*

x206 -1.325e-02 3.911e-03 -3.386 0.000711 \*\*\*

x207 2.348e-01 5.323e-03 44.104 < 2e-16 \*\*\*

x208 6.766e-03 1.383e-03 4.893 1.01e-06 \*\*\*

x209 2.144e-01 7.033e-03 30.480 < 2e-16 \*\*\*

x210 -5.653e-03 1.259e-03 -4.490 7.22e-06 \*\*\*

x211 1.399e-02 2.541e-03 5.505 3.79e-08 \*\*\*

x212 7.126e-03 1.625e-03 4.386 1.17e-05 \*\*\*

x213 4.298e-02 3.315e-03 12.965 < 2e-16 \*\*\*

x214 2.022e-02 3.571e-03 5.662 1.54e-08 \*\*\*

x218 1.995e-01 5.723e-03 34.857 < 2e-16 \*\*\*

x219 1.298e-03 2.725e-04 4.762 1.94e-06 \*\*\*

x220 -1.655e-02 4.152e-03 -3.986 6.78e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1174 on 10038 degrees of freedom

Multiple R-squared: 0.7377, Adjusted R-squared: 0.7372

F-statistic: 1345 on 21 and 10038 DF, p-value: < 2.2e-16

ii.

> probabilities2 = predict(lm.fit3, newdata = jmtest)

> pred2 = rep("0", nrow(jmtest))

> pred2[probabilities2 > 0.5] = "1"

> table(pred2, jmtest$nature)

pred2 0 1

0 7485 133

1 16 434

**MISCLASSIFICATION RATE = 0.076**

(c)

i.

> coef(fit.fwd,21)

(Intercept) x8 x44 x47

-6.095649e-02 -6.145621e-06 5.460678e-03 -6.199597e-08

x56 x63 x69 x78

-1.084668e-02 1.797753e-02 8.363666e-03 -2.638221e-03

x203 x205 x206 x207

2.275970e-01 5.139366e-03 -1.324550e-02 2.347763e-01

x208 x209 x210 x211

6.765808e-03 2.143590e-01 -5.653449e-03 1.398682e-02

x212 x213 x214 x218

7.125842e-03 4.297744e-02 2.021660e-02 1.994953e-01

x219 x220

1.297539e-03 -1.654728e-02

> lm.fit4 = lm(nature~x8+x44+x47+x56+x63+x69+x78+x203+x205+x206+x207+x208+x209

+ +x210+x211+x212+x213+x214+x218+x219+x220, data=jmtrain)

> summary(lm.fit4)

Call:

lm(formula = nature ~ x8 + x44 + x47 + x56 + x63 + x69 + x78 +

x203 + x205 + x206 + x207 + x208 + x209 + x210 + x211 + x212 +

x213 + x214 + x218 + x219 + x220, data = jmtrain)

Residuals:

Min 1Q Median 3Q Max

-1.32843 -0.03658 -0.00875 0.01740 0.97694

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -6.096e-02 7.335e-03 -8.311 < 2e-16 \*\*\*

x8 -6.146e-06 3.308e-07 -18.576 < 2e-16 \*\*\*

x44 5.461e-03 2.836e-04 19.252 < 2e-16 \*\*\*

x47 -6.200e-08 1.464e-08 -4.234 2.32e-05 \*\*\*

x56 -1.085e-02 2.333e-03 -4.649 3.38e-06 \*\*\*

x63 1.798e-02 2.109e-03 8.522 < 2e-16 \*\*\*

x69 8.364e-03 1.364e-03 6.131 9.09e-10 \*\*\*

x78 -2.638e-03 7.952e-04 -3.318 0.000911 \*\*\*

x203 2.276e-01 4.239e-03 53.697 < 2e-16 \*\*\*

x205 5.139e-03 1.659e-03 3.097 0.001960 \*\*

x206 -1.325e-02 3.911e-03 -3.386 0.000711 \*\*\*

x207 2.348e-01 5.323e-03 44.104 < 2e-16 \*\*\*

x208 6.766e-03 1.383e-03 4.893 1.01e-06 \*\*\*

x209 2.144e-01 7.033e-03 30.480 < 2e-16 \*\*\*

x210 -5.653e-03 1.259e-03 -4.490 7.22e-06 \*\*\*

x211 1.399e-02 2.541e-03 5.505 3.79e-08 \*\*\*

x212 7.126e-03 1.625e-03 4.386 1.17e-05 \*\*\*

x213 4.298e-02 3.315e-03 12.965 < 2e-16 \*\*\*

x214 2.022e-02 3.571e-03 5.662 1.54e-08 \*\*\*

x218 1.995e-01 5.723e-03 34.857 < 2e-16 \*\*\*

x219 1.298e-03 2.725e-04 4.762 1.94e-06 \*\*\*

x220 -1.655e-02 4.152e-03 -3.986 6.78e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1174 on 10038 degrees of freedom

Multiple R-squared: 0.7377, Adjusted R-squared: 0.7372

F-statistic: 1345 on 21 and 10038 DF, p-value: < 2.2e-16

ii.

> probabilities3 = predict(lm.fit4, newdata = jmtest)

> pred3 = rep("0", nrow(jmtest))

> pred3[probabilities3 > 0.5] = "1"

> table(pred3, jmtest$nature)

pred3 0 1

0 7485 133

1 16 434

**MISCLASSIFICATION RATE = 0.018%**

**After the selecting 21 variables model for both the forward and backward selection of models. These criteria were done by the basis of parsimony. This helps in the easy interpretation of our results in our model. Adding more variables will make our justification easy. Hence, BIC will be the preferred model of selection.**

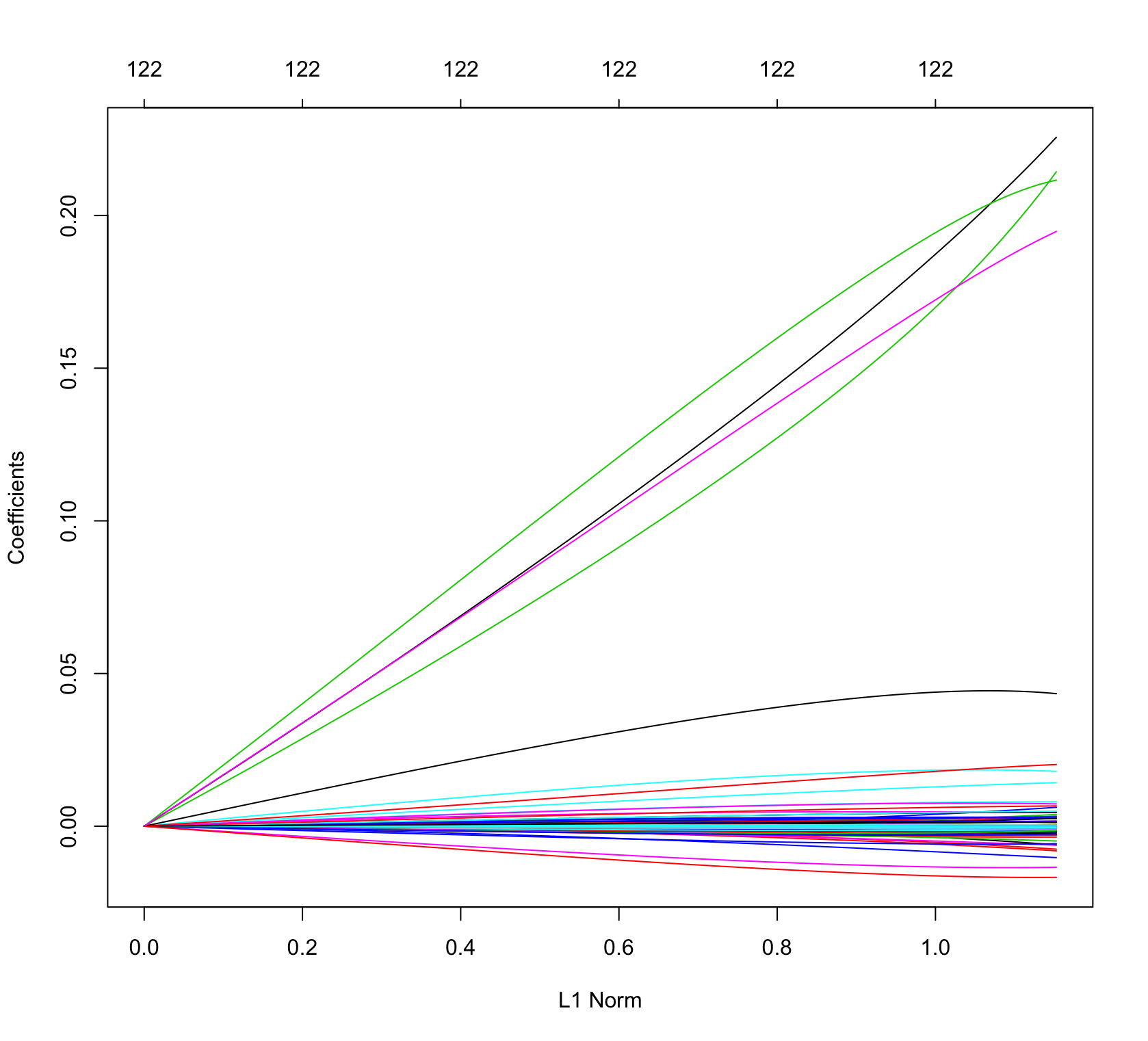
**Question 3.**

**i.**

> #Ridge Regression#

> ridge.fit = glmnet(xtrain,ytrain, alpha=0)

> plot(ridge.fit)



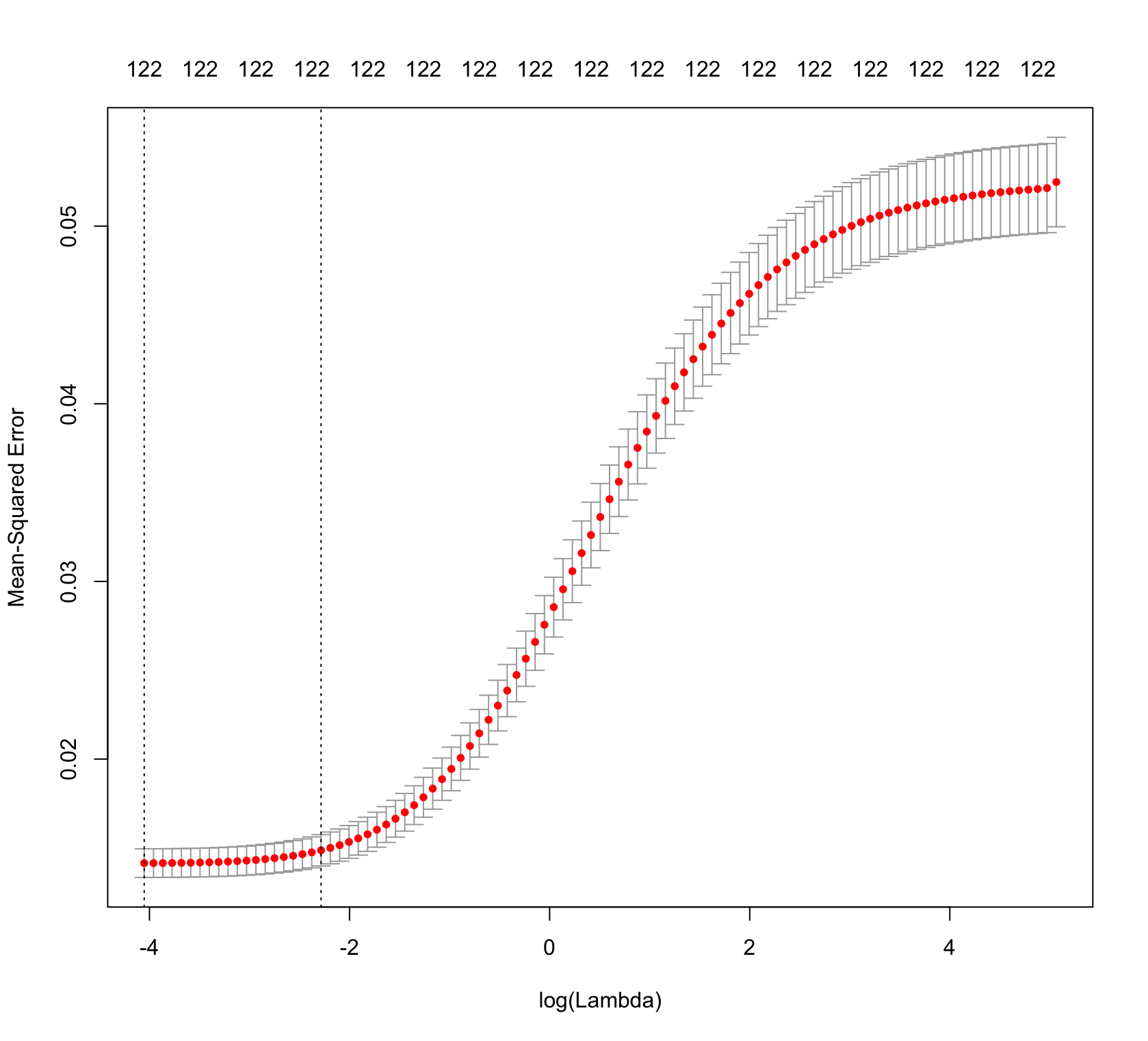
ii.

> set.seed(8)

> cv.out=cv.glmnet(xtrain, ytrain, alpha=0)

> plot(cv.out)

> best.lambda=cv.out$lambda.min



iii.

> #Prediction#

> pred.ridge= predict(ridge.fit, s=best.lambda, newx=xtest)

> mean((ridge.pred-ytest)^2)

[1] 0.01710797

iii.

> #Ridge confusion matrix#

> matrix.ridge = rep("0", nrow(jmtest))

> matrix.ridge[pred.ridge > 0.5] = "1"

> table(pred.ridge, jmtest$nature)

pred.ridge 0 1

-0.126004821102568 1 0

-0.125526986577561 1 0

-0.118575060506295 1 0

-0.107151218412276 1 0

-0.102834139826285 1 0

-0.102481457270084 1 0

-0.0947964821216436 1 0

-0.0942224742408934 1 0

-0.0925634364003604 1 0

-0.0922392683183185 1 0

-0.0911829691753026 1 0

-0.0910859482575885 1 0

-0.0906760570648758 1 0

-0.0906124719967947 1 0

-0.0900587281414979 1 0

-0.0900253064274502 1 0

-0.0896485897858907 1 0

-0.0895229516860182 1 0

-0.0887237570972074 1 0

-0.0885823178614114 1 0

-0.0874174212943559 1 0

-0.0870159474454306 1 0

-0.0863322750917963 1 0

-0.08576389865328 1 0

-0.0852558735452552 1 0

-0.0840290677691018 1 0

-0.0838898242479679 1 0

-0.0823177758031232 1 0

-0.0821524494662132 1 0

-0.0815707782324043 1 0

-0.0814078376956621 1 0

-0.0805773137936298 1 0

-0.0802375226322408 1 0

-0.0798984812448769 1 0

-0.07987931275975 1 0

-0.0797722221210207 1 0

-0.0795300135986662 1 0

-0.0794224710754155 1 0

-0.0793604632669856 1 0

-0.0790941863883919 1 0

-0.0790558421263186 1 0

-0.0788921281729643 1 0

-0.0786769193535357 1 0

-0.0784573693699976 1 0

-0.0781540887683011 1 0

-0.0779217670457913 1 0

-0.0772260153435154 1 0

-0.0770187265138966 1 0

-0.0766927132686927 1 0

-0.0764760845903788 1 0

-0.0762002284763761 1 0

-0.0759382573046726 1 0

-0.07570938703035 1 0

-0.0756817699696447 1 0

-0.0756555491960101 1 0

-0.0750726888560602 1 0

-0.0750171321951754 1 0

-0.0746840148457263 1 0

-0.074537389034107 1 0

-0.0742145408186018 1 0

-0.0741340101710362 1 0

-0.0741129377378831 1 0

-0.073909409157447 1 0

-0.0735312492662481 1 0

-0.0734395492440715 1 0

-0.072499373216192 1 0

-0.0719161098464952 1 0

-0.071869411337629 1 0

-0.0716197662994166 1 0

-0.0715998385021321 1 0

-0.0714665485040227 1 0

-0.0713393672401651 1 0

-0.0708618289404992 1 0

-0.0708115370688936 1 0

-0.0707371318159836 1 0

-0.0706098105247753 1 0

-0.0701954884540723 1 0

-0.0701105416209056 1 0

-0.07006255643798 1 0

-0.0699430339375304 1 0

-0.0696410351963949 1 0

-0.0695379801921293 1 0

-0.0694872473360501 1 0

-0.0692216168328937 1 0

-0.0691635298125461 1 0

-0.0689047590362474 1 0

-0.0688589214687656 1 0

-0.068596183138156 1 0

-0.0684350767111569 1 0

-0.0683928149088542 1 0

-0.0683244527098589 1 0

-0.067831549918163 1 0

-0.0674941385079702 1 0

-0.0674767648253869 1 0

-0.0672660110494242 1 0

-0.0671221047444192 1 0

-0.0670823094212598 1 0

-0.0667194774350354 1 0

-0.0665959032062855 1 0

-0.0663988084766542 1 0

-0.0663769289795792 1 0

-0.0663492984357661 1 0

-0.0657126118542945 1 0

-0.0656902447801979 1 0

-0.0653952729470921 1 0

-0.0653588747565979 1 0

-0.0651404427002531 1 0

-0.0650643211873501 1 0

-0.0650423867522077 1 0

-0.0649238465424805 1 0

-0.0647126458617569 1 0

-0.0646588186402283 1 0

-0.064429572646099 1 0

-0.0643138182861023 1 0

-0.0641996316391891 1 0

-0.0641765262665724 1 0

-0.0641528926618461 1 0

-0.064127166079796 1 0

-0.0640810645599751 1 0

-0.0640496049611504 1 0

-0.0640111892125667 1 0

-0.0639400749979848 1 0

-0.0638882051922535 1 0

-0.0638660642186132 1 0

-0.0638612944820164 1 0

-0.0637754335660902 1 0

-0.0636458357556611 1 0

-0.0635540138562916 1 0

-0.0632401251643246 1 0

-0.0632382741176979 1 0

-0.0629860300004133 1 0

-0.0628649601348991 1 0

-0.0627503344993977 1 0

-0.0626681632256319 1 0

-0.0625939888487758 1 0

-0.0625279704877179 1 0

-0.0624351684154414 1 0

-0.0624189496028021 1 0

-0.0624109083653805 1 0

-0.0623569323502147 1 0

-0.0623410982115466 1 0

-0.0622133875175713 1 0

-0.0621550675344062 1 0

-0.0620912047006936 1 0

-0.0620799821169824 1 0

-0.0619824678694432 1 0

-0.0619479973770446 1 0

-0.0618873972181124 1 0

-0.061773522805694 1 0

-0.0615844791160418 1 0

-0.0614457602260021 1 0

-0.0613949758987725 1 0

-0.0613208862750165 1 0

-0.0613084047634742 1 0

-0.0612892772438957 1 0

-0.060857162189944 1 0

-0.0607923083394641 1 0

-0.0607494961028499 1 0

-0.060594778944352 1 0

-0.0605223948944703 1 0

-0.060436266739595 1 0

-0.0603690285359075 1 0

-0.0603174221429492 1 0

-0.0602433789266288 1 0

-0.0600411752748888 1 0

-0.060032789585928 1 0

-0.0599728088915153 1 0

-0.0599363888753379 1 0

-0.0597942525274734 1 0

-0.0597798391251236 1 0

-0.0597797909037893 1 0

-0.0597472368153227 1 0

-0.059639842890629 1 0

-0.0595934087358271 1 0

-0.0595904532483936 1 0

-0.0594520999114915 1 0

-0.059236834751798 1 0

-0.0592126767681905 1 0

-0.0592035868952659 1 0

-0.0590887991252746 1 0

-0.0590033822574823 1 0

-0.0589911262494372 1 0

-0.0589591834899691 1 0

-0.0589269378679924 1 0

-0.0588049173995296 1 0

-0.0587999474763936 1 0

-0.058679688861246 1 0

-0.0585032157608453 1 0

-0.0584934177705653 1 0

-0.0583287126758623 1 0

-0.0583200702076411 1 0

-0.0582431225660366 1 0

-0.0581654696922278 1 0

-0.0580937021366643 1 0

-0.0580223975711583 1 0

-0.0580154620696984 1 0

-0.0579033263347667 1 0

-0.0577525706418898 1 0

-0.05763500944356 1 0

-0.0574398153475271 1 0

-0.0571845546020277 1 0

-0.0571265461330299 1 0

-0.0570146970568229 1 0

-0.056924419782904 1 0

-0.056902607968575 1 0

-0.0567913688727086 1 0

-0.0567317182918215 1 0

-0.0565873959132215 1 0

-0.0565732772304752 1 0

-0.0565656589623315 1 0

-0.0564277738967253 1 0

-0.0563424954262288 1 0

-0.0563364550350087 1 0

-0.0563312383816267 1 0

-0.0562687801384538 1 0

-0.0562293667210735 1 0

-0.0560239510244507 1 0

-0.0560197340729062 1 0

-0.0559644683173547 1 0

-0.0559353973809046 1 0

-0.0558942542733661 1 0

-0.0558886538981142 1 0

-0.0558831343696254 1 0

-0.0557633530812715 1 0

-0.0557460737280674 1 0

-0.0556373016941373 1 0

-0.0554881603245907 1 0

-0.0554255136467302 1 0

-0.055380430694863 1 0

-0.0553641617557988 1 0

-0.0553512739872731 1 0

-0.0553415196896768 1 0

-0.0553059171911552 1 0

-0.0552731638525465 1 0

-0.0551582519084782 1 0

-0.0550502112512394 1 0

-0.0549899622959297 1 0

-0.0549383353158137 1 0

-0.0549059459106881 1 0

-0.0548086802127085 1 0

-0.0547839584312364 1 0

-0.0544454108854271 1 0

-0.0543728329411984 1 0

-0.0543548891350527 1 0

-0.0543218121755442 1 0

-0.0542640637467237 1 0

-0.0541619774723312 1 0

-0.0541499774152007 1 0

-0.0541348712174132 1 0

-0.0540889728612907 1 0

-0.0540635903202271 1 0

-0.0538902020340912 1 0

-0.0536298707770293 1 0

-0.0535221371917886 1 0

-0.0534716895100556 1 0

-0.0533616099280158 1 0

-0.0533099578601223 1 0

-0.0532886390545655 1 0

-0.0532731092887796 1 0

-0.0532089739427232 1 0

-0.0530619087339182 1 0

-0.0529380530642609 1 0

-0.0528247114350229 1 0

-0.0527749854872009 1 0

-0.0527206052399403 1 0

-0.0527069104884477 1 0

-0.052638271885513 1 0

-0.0524681077163868 1 0

-0.0524136462898311 1 0

-0.0523972985682505 1 0

-0.0523965604674886 1 0

-0.05238095138506 1 0

-0.0523779759483348 1 0

-0.0523345316831195 1 0

-0.0522584045748968 1 0

-0.052226073468491 1 0

-0.0522072029242085 1 0

-0.0521245088004272 1 0

-0.0521128295568698 1 0

-0.0521002602030021 1 0

-0.0520635187437896 1 0

-0.0519429235221642 1 0

-0.0518713341152683 1 0

-0.0518234750280313 1 0

-0.0518199235822474 1 0

-0.0517483309855009 1 0

-0.0517329755199826 1 0

-0.0516383651907601 1 0

-0.0516162416967842 1 0

-0.0516089244094782 1 0

-0.051505102077988 1 0

-0.0514943649639421 1 0

-0.0514001574649291 1 0

-0.0513416768578472 1 0

-0.0513410030832121 1 0

-0.0512811229828989 1 0

-0.0512782597533414 1 0

-0.0512548675075899 1 0

-0.0510972160218585 1 0

-0.0510615479068186 1 0

-0.051032963026875 1 0

-0.0509873198413459 1 0

-0.0506764273783843 1 0

-0.0504650664201699 1 0

-0.0504076422783461 1 0

-0.0503967850120829 1 0

-0.0503371107869211 1 0

-0.0501862844610701 1 0

-0.0501348391938893 1 0

-0.0501172061386255 1 0

-0.0500945186207639 1 0

-0.0500791457086636 1 0

-0.0500234105023567 1 0

-0.0499734203960069 1 0

-0.0499634116092828 1 0

-0.0499533621671434 1 0

-0.049848794706457 1 0

-0.0497598195020591 1 0

-0.049706994499556 1 0

-0.0496747656200128 1 0

-0.0496565158273085 1 0

-0.0496361523467247 1 0

-0.0494491033672908 1 0

-0.0494025259378149 1 0

-0.0493291076648826 1 0

-0.0492362321317983 1 0

-0.0491470746029868 1 0

-0.0491355460906821 1 0

-0.0491279201050681 1 0

-0.0491068118628039 1 0

-0.0490337712450502 1 0

-0.0490033117371334 1 0

-0.0490002600207408 1 0

-0.0489404673553741 1 0

-0.0486700666958374 1 0

-0.048665172263873 1 0

-0.048599042754791 1 0

-0.0485314272787684 1 0

-0.0485067943339217 1 0

-0.0484671471135858 1 0

-0.0483843912604354 1 0

-0.048329172026542 1 0

-0.0482179198640936 1 0

-0.048202062100126 1 0

-0.0481547133362255 1 0

-0.0481381509394949 1 0

-0.0479965873996699 1 0

-0.0479656964442491 1 0

-0.0479289701198056 1 0

-0.0477850465890733 1 0

-0.0477713722342541 1 0

-0.047756667520266 1 0

-0.0476923935216174 1 0

-0.047682679043289 1 0

-0.0476176996855672 1 0

-0.0476141611263461 1 0

-0.0476032733600343 1 0

-0.0475489096213788 1 0

-0.0475268215849635 1 0

-0.0474588711693678 1 0

-0.0474571080135222 1 0

-0.0474242143157084 1 0

-0.0474106211661703 1 0

-0.047407109716071 1 0

-0.0473908069134388 1 0

-0.0473501013676766 1 0

-0.0472641129324232 1 0

-0.047260824291521 1 0

-0.0472372185142147 1 0

-0.0472053605590406 1 0

-0.047161616790666 1 0

-0.0471319179071574 1 0

-0.0471181818835061 1 0

-0.0470548665049104 1 0

-0.0469725194100064 1 0

-0.0469434940841328 1 0

-0.0468492769346354 1 0

-0.0468403796882706 1 0

-0.0468113289296379 1 0

-0.0467719873502261 1 0

-0.0467430035485273 1 0

-0.0467267887502643 1 0

-0.0467028947263545 1 0

-0.0466415504667365 1 0

-0.0466355875514654 1 0

-0.0466008571435134 1 0

-0.0465791335200631 1 0

-0.0465689548126807 1 0

-0.0465543196754364 1 0

-0.0465308653941355 1 0

-0.0463828049741653 1 0

-0.0463593622172349 1 0

-0.0463446582247859 1 0

-0.0463148870439238 1 0

-0.0461696639726324 1 0

-0.0460950688141936 1 0

-0.0460399024927814 1 0

-0.0460179791717329 1 0

-0.0460081792191658 1 0

-0.0458988735858672 1 0

-0.0458752500907776 1 0

-0.0458315923272997 1 0

-0.0458304537654835 1 0

-0.0458137044952113 1 0

-0.0457618422925715 1 0

-0.0457128809755262 1 0

-0.0456792841413582 1 0

-0.0456455936668179 1 0

-0.0456152545701192 1 0

-0.0456080780927842 1 0

-0.0455491149897688 1 0

-0.0455004007463573 1 0

-0.0454975556584557 1 0

-0.0454730762336756 1 0

-0.0454616748354188 1 0

-0.0454118552194767 1 0

-0.0453968884097071 1 0

-0.0452720003129673 1 0

-0.0452483459466427 1 0

-0.045243962875151 1 0

-0.0452354647721832 1 0

-0.0452280071837532 1 0

-0.0450849353837283 1 0

-0.0450786204170959 1 0

-0.0450776147646297 1 0

-0.0449948649663636 1 0

-0.0449465887594126 1 0

-0.04491473665499 1 0

-0.044868202593357 1 0

-0.0447604376616222 1 0

-0.0447530003843039 1 0

-0.0447168513589213 1 0

-0.0447022906755847 1 0

-0.0447006456121992 1 0

-0.0446387432980853 1 0

-0.0446201724163094 1 0

-0.0445380876305488 1 0

-0.0445348736630743 1 0

-0.0445064392035045 1 0

-0.0445010670290711 1 0

-0.0443543899471207 1 0

-0.0443348899426664 1 0

-0.0442497744049991 1 0

-0.0442113419339244 1 0

-0.0442111380945193 1 0

-0.0442043476840392 1 0

-0.044164342892487 1 0

-0.0441494204307159 1 0

-0.0440933032710241 1 0

-0.0440880344590661 1 0

-0.0440711054803366 1 0

-0.044068766991332 1 0

-0.0440576760534861 1 0

-0.0439950427457694 1 0

-0.0439790937061129 1 0

-0.0439446011670754 1 0

-0.0438479267091724 1 0

-0.0438124661575514 1 0

-0.0437728594566951 1 0

-0.0437637608690965 1 0

-0.0437509688158266 1 0

-0.0436498000789151 1 0

-0.043644811248923 1 0

-0.0436328413187174 1 0

-0.043604180467921 1 0

-0.043549381487087 1 0

-0.0434585891475514 1 0

-0.0434361266697868 1 0

-0.0434250958014804 1 0

-0.0433359149457693 1 0

-0.0433293449356477 1 0

-0.0433217713817729 1 0

-0.043310849986758 1 0

-0.0431622040354623 1 0

-0.0430034091325919 1 0

-0.0430003448589508 1 0

-0.0429503956605623 1 0

-0.0429139038650433 1 0

-0.0428811199548836 1 0

-0.0428615071411846 1 0

-0.0428406455213747 1 0

-0.042833135097153 1 0

-0.0428150424196294 1 0

-0.0428119082720056 1 0

-0.0427356148557358 1 0

-0.0427312014638574 1 0

-0.042689269800304 1 0

-0.0426325445384302 1 0

-0.0425490000433945 1 0

-0.0424743871812506 1 0

-0.042385488014554 1 0

-0.0423631473035906 1 0

-0.0423628498186022 1 0

-0.0423581133470087 1 0

-0.0423368536337294 1 0

-0.0423013240585718 1 0

-0.0422968304150531 1 0

-0.0422379135561671 1 0

-0.0422374378900624 1 0

-0.0422362552966077 1 0

[ reached getOption("max.print") -- omitted 7568 rows ]

**iv.**

**CALCULATION OF FP, FN, SENSITITVITY, SPECIFICITY FOR RIDGE MODEL**

**FP rate =**

**FN rate =**

**SPECIFCITY =**

**SENSITIVITY =**

**MISCLASSIFICATION RATE = 0.034%**

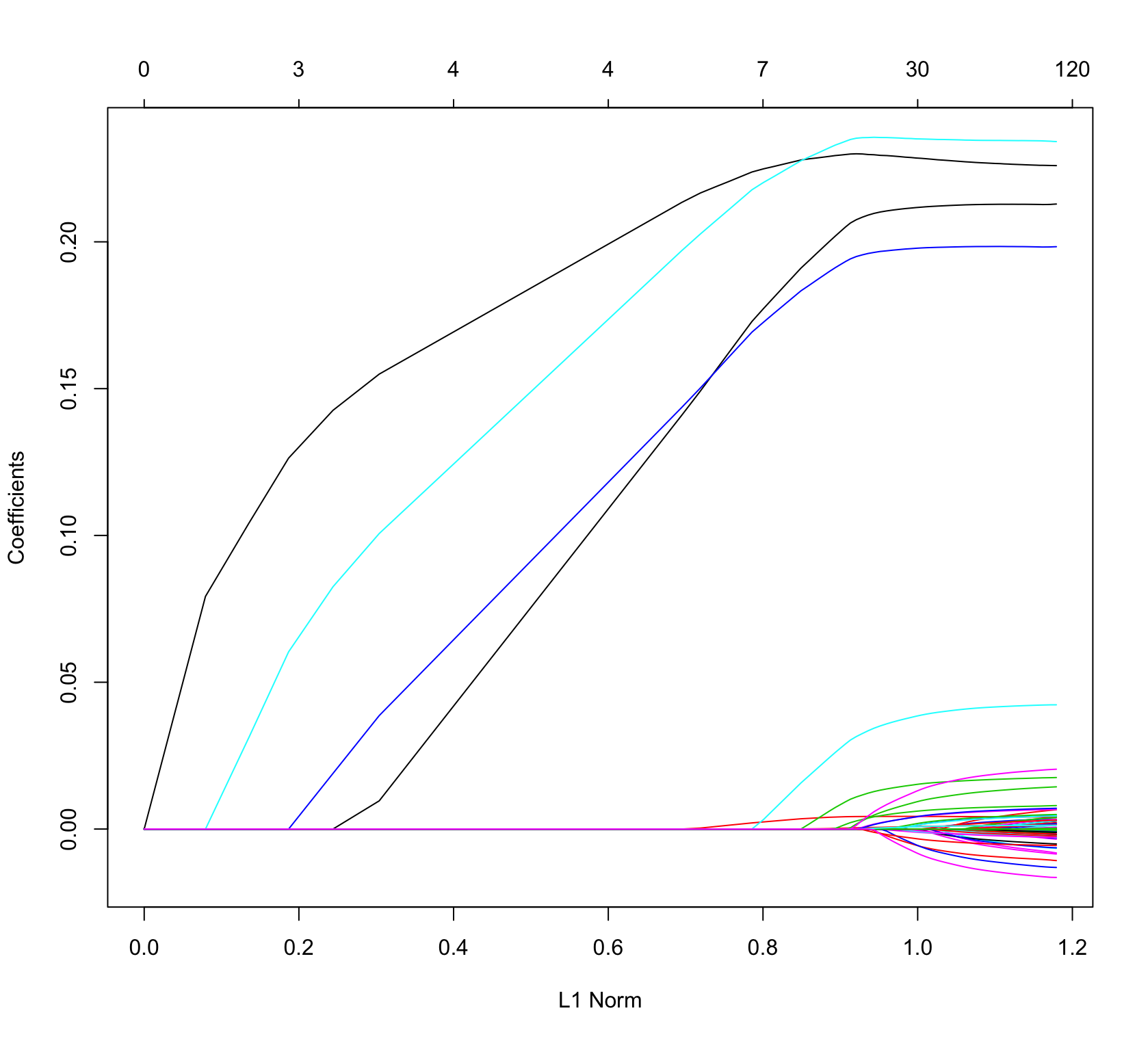
(b)

i.

> #Lasso Regression#

> lasso.fit = glmnet(xtrain,ytrain, alpha=1)

> plot(lasso.fit)



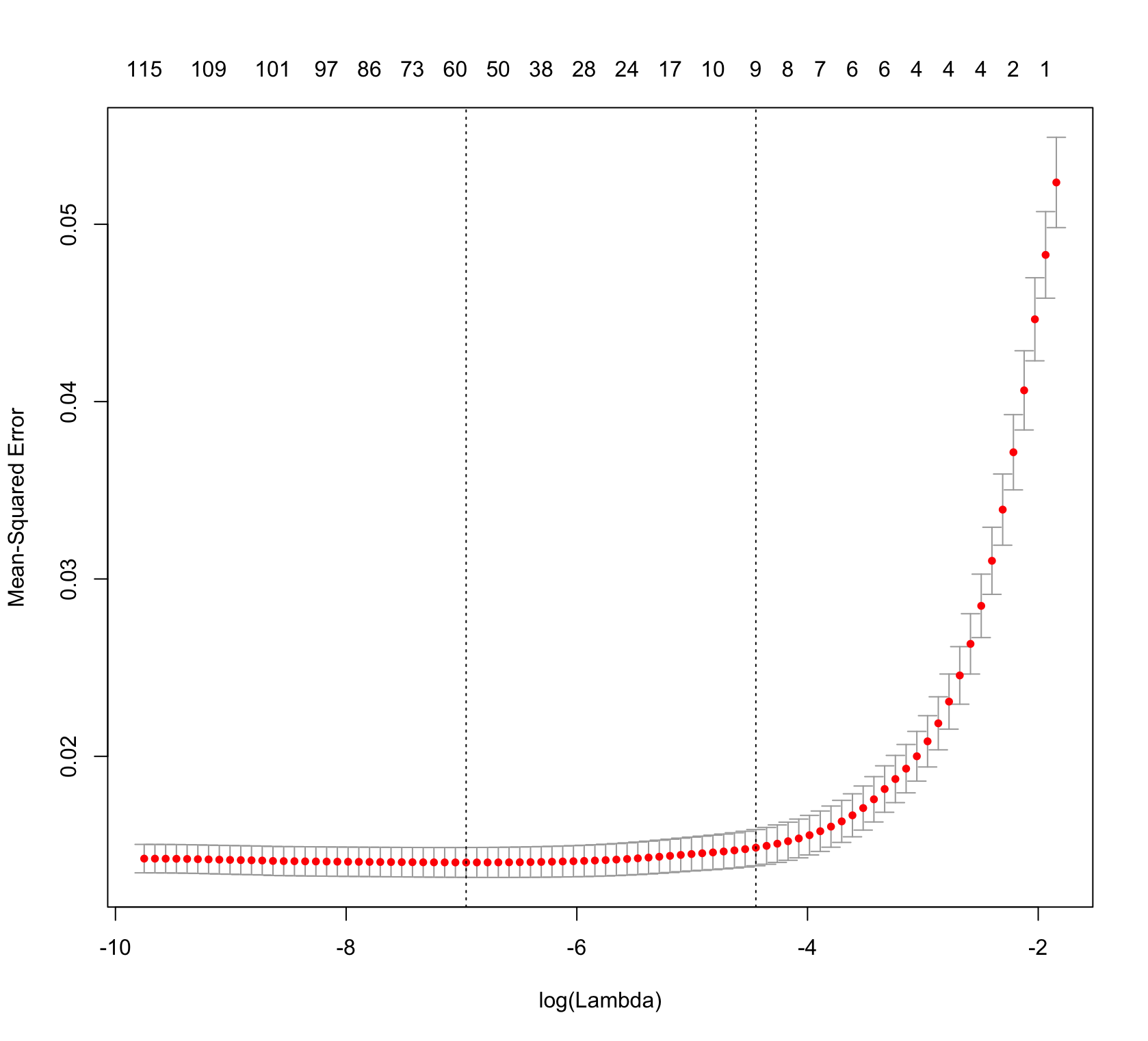
ii.

> set.seed(8)

> cv.out2=cv.glmnet(xtrain, ytrain, alpha=1)

> plot(cv.out2)

> best.lambda2=cv.out$lambda.mi



iii.

> ##Prediction#

> pred.lasso= predict(lasso.fit, s=best.lambda2, newx=xtest)

> mean((pred.lasso-ytest)^2)

[1] 0.01919147

iii.

> #Lasso Drop#

> drop.lasso = glmnet(xtrain, ytrain, alpha = 1)

> coeff.lasso = predict(drop.lasso, type="coefficients", s=best.lambda2)

> coeff.lasso

123 x 1 sparse Matrix of class "dgCMatrix"

1

(Intercept) -1.501326e-02

x1 .

x2 .

x3 .

x4 .

x5 .

x6 .

x7 .

x8 -3.402006e-06

x9 .

x10 .

x11 .

x12 .

x13 .

x14 .

x15 .

x16 .

x17 .

x18 .

x19 .

x42 .

x43 .

x44 3.210910e-03

x45 .

x46 .

x47 .

x48 .

x49 .

x50 .

x51 .

x52 .

x53 .

x54 .

x55 .

x56 .

x57 .

x58 .

x59 .

x60 .

x61 .

x62 1.305752e-06

x63 .

x64 .

x65 .

x66 .

x67 .

x68 .

x69 .

x70 .

x71 .

x72 .

x73 .

x74 .

x75 .

x76 .

x77 .

x78 .

x79 .

x80 .

x81 .

x82 .

x83 .

x84 .

x85 .

x86 .

x87 .

x88 .

x89 .

x90 .

x91 .

x92 .

x93 .

x94 .

x95 .

x96 .

x97 .

x98 .

x99 .

x100 .

x101 .

x102 .

x103 .

x104 .

x105 .

x106 .

x107 .

x108 .

x109 .

x110 .

x111 .

x112 .

x113 .

x114 .

x115 .

x116 .

x117 .

x118 .

x119 .

x120 .

x121 .

x122 .

x123 .

x124 .

x201 .

x202 .

x203 2.270440e-01

x204 .

x205 .

x206 .

x207 2.255851e-01

x208 .

x209 1.872810e-01

x210 .

x211 .

x212 .

x213 1.229261e-02

x214 .

x215 .

x216 .

x217 .

x218 1.804008e-01

x219 .

x220 .

iv.

#Lasso confusion matrix#

matrix.lasso = rep("0", nrow(jmtest))

matrix.lasso[pred.lasso > 0.5] = "1"

table(pred.lasso, jmtest$nature)

pred.lasso 0 1

-0.0294197840074434 1 0

-0.0274104267590063 1 0

-0.0269177945221055 1 0

-0.0268038150894878 1 0

-0.0260793914133251 1 0)

-0.0255966655390437 1 0

-0.0240700233801563 1 0

-0.0232128839017063 1 0

-0.023106552556712 1 0

-0.0223303989422927 1 0

-0.0219298744686195 1 0

-0.0209224528819041 1 0

-0.0207142817581363 1 0

-0.0205283899114665 1 0

-0.0204407040750323 1 0

-0.0203190852241952 1 0

-0.0201960701038604 1 0

-0.0200552624166176 1 0

-0.0198124718668363 1 0

-0.0196981461100552 1 0

-0.0196532351516916 1 0

-0.0196069845462527 1 0

-0.0194058661438804 1 0

-0.0191891661142746 1 0

-0.0190976655349106 1 0

-0.0189484656787299 1 0

-0.0188901052982306 1 0

-0.0187323801919915 1 0

-0.018580121999379 1 0

-0.0183467131212777 1 0

-0.0182317160119762 1 0

-0.0181289796992191 1 0

-0.0179571948510515 1 0

-0.0177516659870185 1 0

-0.0177407448061483 1 0

-0.0176908240819965 1 0

-0.0174327417405938 1 0

-0.0173261760991042 1 0

-0.0172502392531728 1 0

-0.017218638641462 1 0

-0.0170745400298758 1 0

-0.0169574757213438 1 0

-0.016921347306343 1 0

-0.0169142844377959 1 0

-0.016827759193679 1 0

-0.0167887453344126 1 0

-0.0167791333431186 1 0

-0.0167628587939013 1 0

-0.016681068277598 1 0

-0.0166388309553837 1 0

-0.0166090087753377 1 0

-0.0163036066171385 1 0

-0.0161816161418887 1 0

-0.0161619458200139 1 0

-0.0160863744787246 1 0

-0.0159875014483098 1 0

-0.0159033685923313 1 0

-0.0158569111800985 1 0

-0.0158138724080845 1 0

-0.0155422770384628 1 0

-0.0153428420149723 1 0

-0.0152709808106528 1 0

-0.0152284354272881 1 0

-0.0151250314690741 1 0

-0.0150567316836806 1 0

-0.0149867514719434 1 0

-0.0148913109282841 1 0

-0.0148320183707067 1 0

-0.0148101146597604 1 0

-0.0147933107156608 1 0

-0.0147471816635055 1 0

-0.0147464447653845 1 0

-0.0147381905420609 1 0

-0.0146935442558885 1 0

-0.0146456988919412 1 0

-0.0146364186627303 1 0

-0.0146361690923526 1 0

-0.0145592258353375 1 0

-0.0144506141848984 1 0

-0.0143883633687169 1 0

-0.0143732659523962 1 0

-0.0143365852473901 1 0

-0.0142936946384273 1 0

-0.0142248598230885 1 0

-0.0141221390113083 1 0

-0.0139881318352818 1 0

-0.0139753877971421 1 0

-0.0138673310527875 1 0

-0.0138277290494685 1 0

-0.0137795861078836 1 0

-0.0137670154979018 1 0

-0.013761685673678 1 0

-0.0136904438728803 1 0

-0.0135856065870306 1 0

-0.013474364674142 1 0

-0.0134232120562246 1 0

-0.0133869178855907 1 0

-0.0133648042433413 1 0

-0.0132944162091041 1 0

-0.0132756158703103 1 0

-0.0132210215060981 1 0

-0.013214547728081 1 0

-0.0131959852284558 1 0

-0.0131910611656456 1 0

-0.0131117044593609 1 0

-0.0130591292481375 1 0

-0.0130106099468939 1 0

-0.012934973333661 1 0

-0.0129029971015574 1 0

-0.0128692588576813 1 0

-0.0128534211611171 1 0

-0.0128175722655629 1 0

-0.0128058467726207 1 0

-0.0127738017221586 1 0

-0.0127703948872731 1 0

-0.0127422955065519 1 0

-0.0127009422914517 1 0

-0.012635520024025 1 0

-0.0125161415696426 1 0

-0.0124883881453226 1 0

-0.0124517704982037 1 0

-0.0123940218923069 1 0

-0.0123154469317515 1 0

-0.0122197314946249 1 0

-0.0121548815576353 1 0

-0.012118576199365 1 0

-0.0120733599014552 1 0

-0.012057340400967 1 0

-0.0120238023471044 1 0

-0.0120197118383541 1 0

-0.0120094059434565 1 0

-0.0119118576929785 1 0

-0.0119048550009475 1 0

-0.0118857139570095 1 0

-0.0118583281175167 1 0

-0.0118045475481709 1 0

-0.0117652201256803 1 0

-0.0117343168899614 1 0

-0.0117305192273442 1 0

-0.0117064728981699 1 0

-0.0116484984656491 1 0

-0.0116472187794158 1 0

-0.0116359456829246 1 0

-0.0116329274938996 1 0

-0.0116126266084283 1 0

-0.0115253749581662 1 0

-0.0114552056414839 1 0

-0.0114537841743048 1 0

-0.0114449402190319 1 0

-0.011440385688147 1 0

-0.0113654194619562 1 0

-0.011341507445457 1 0

-0.0113293964777261 1 0

-0.0113193019100763 1 0

-0.0113182139525457 1 0

-0.0112117527098049 1 0

-0.0111902233601352 1 0

-0.0111485950206339 1 0

-0.0111284433425377 1 0

-0.0110826715361061 1 0

-0.0110409725365757 1 0

-0.0110187269240732 1 0

-0.0110145406545772 1 0

-0.0109592643640658 1 0

-0.0109302906403607 1 0

-0.0108884405956239 1 0

-0.0108797049219737 1 0

-0.0108257014234186 1 0

-0.0108136061714567 1 0

-0.0107689031031928 1 0

-0.0107449996933935 1 0

-0.0107435086684921 1 0

-0.0107401158759345 1 0

-0.0107328672189033 1 0

-0.0107112516968076 1 0

-0.0106410248510893 1 0

-0.01063037132251 1 0

-0.010607107700704 1 0

-0.0105561480392046 1 0

-0.0105477161684879 1 0

-0.0105452084875044 1 0

-0.0105442183280836 1 0

-0.0105249681149105 1 0

-0.0105183489344285 1 0

-0.0105152009218791 1 0

-0.0104954366305273 1 0

-0.0104615228194433 1 0

-0.0104444438678882 1 0

-0.0104248702532365 1 0

-0.0103476639936319 1 0

-0.0102917538353651 1 0

-0.0101525465142298 1 0

-0.0101478203854374 1 0

-0.010145095617519 1 0

-0.0101091205664244 1 0

-0.0100666147518492 1 0

-0.0100568768160654 1 0

-0.0100093780499385 1 0

-0.00998948438069035 1 0

-0.00997447313295642 1 0

-0.00995500539864047 1 0

-0.00994792170373519 1 0

-0.00991616670367728 1 0

-0.00990034206966562 1 0

-0.00989330142157677 1 0

-0.00989195544751957 1 0

-0.00986622791749525 1 0

-0.00985054437076358 1 0

-0.00984312296223408 1 0

-0.00980522713881481 1 0

-0.00979917585090317 1 0

-0.0097748614047109 1 0

-0.00975174849175002 1 0

-0.00974213948813666 1 0

-0.0096540300041668 1 0

-0.00963767176561334 1 0

-0.00962740145110596 1 0

-0.00959023898697368 1 0

-0.00956247723345023 1 0

-0.00956157978439029 1 0

-0.00947949581567723 1 0

-0.00946971390375621 1 0

-0.00946238223020914 1 0

-0.00945063922648781 1 0

-0.00940365416513893 1 0

-0.00939602306187856 1 0

-0.00935679228390677 1 0

-0.00931196695439659 1 0

-0.00929245903359418 1 0

-0.00928113208372951 1 0

-0.00927545322899735 1 0

-0.00924913092671962 1 0

-0.00923975322434008 1 0

-0.00920536275329895 1 0

-0.00919172952117843 1 0

-0.00915881261082222 1 0

-0.00912092967604213 1 0

-0.00912053926263534 1 0

-0.0091152645937223 1 0

-0.00907440375050887 1 0

-0.00906462134629928 1 0

-0.00901356704634187 1 0

-0.00898422838912893 1 0

-0.00893970086051136 1 0

-0.00893867485078474 1 0

-0.00890894462657413 1 0

-0.00884632669501357 1 0

-0.00883416140071942 1 0

-0.0087791461342983 1 0

-0.00875165820323852 1 0

-0.00875009434027626 1 0

-0.00873320772744217 1 0

-0.00872696991561988 1 0

-0.00872555744190389 1 0

-0.00870550853951735 1 0

-0.00869589749470067 1 0

-0.00867499340224718 1 0

-0.00867159873287638 1 0

-0.00864832259019334 1 0

-0.00864296845117934 1 0

-0.00856844135546253 1 0

-0.00855495033665121 1 0

-0.00853097258732082 1 0

-0.00852150548632007 1 0

-0.00851380395206725 1 0

-0.00849724266034609 1 0

-0.00847706862686007 1 0

-0.00845673007697405 1 0

-0.00845219690843895 1 0

-0.00843961571325815 1 0

-0.00839292366432949 1 0

-0.00837341298678013 1 0

-0.00836183875767582 1 0

-0.00830378441700676 1 0

-0.00821378245622307 1 0

-0.00821077977201417 1 0

-0.00819030735508113 1 0

-0.00818648214389772 1 0

-0.00815453877441923 1 0

-0.00814287291825986 1 0

-0.00813514403681808 1 0

-0.00813282774997535 1 0

-0.00811515239554686 1 0

-0.00810732546411836 1 0

-0.00806029484151469 1 0

-0.00805724097425277 1 0

-0.00804951263828058 1 0

-0.0080383355876123 1 0

-0.00802733478839237 1 0

-0.00801875088317083 1 0

-0.00800422625905685 1 0

-0.00798198728617519 1 0

-0.00797198157357402 1 0

-0.00796512629390794 1 0

-0.00795938969150822 1 0

-0.00788275207141862 1 0

-0.00785246228758266 1 0

-0.0078487853629617 1 0

-0.0078029729263125 1 0

-0.00779437453313502 1 0

-0.00779206082917117 1 0

-0.00777434011870444 1 0

-0.00773282990675887 1 0

-0.00772084668246498 1 0

-0.00767946731090394 1 0

-0.00763748770862116 1 0

-0.00762409647220031 1 0

-0.00762290107015975 1 0

-0.00762176729098205 1 0

-0.00762092576022124 1 0

-0.00760228607405231 1 0

-0.00760129721272951 1 0

-0.00759410635464991 1 0

-0.00758747918516499 1 0

-0.00756462383158471 1 0

-0.00754518215204352 1 0

-0.00753119361745824 1 0

-0.00752218983446874 1 0

-0.00751676173510611 1 0

-0.00743576546168671 1 0

-0.007426638182057 1 0

-0.00735192115848168 1 0

-0.00731060434484064 1 0

-0.007289126298911 1 0

-0.0072797325681236 1 0

-0.00726999648159128 1 0

-0.00724709219799617 1 0

-0.0072416735501247 1 0

-0.00715196988252217 1 0

-0.00715141760019543 1 0

-0.00714782127744099 1 0

-0.00708147188938192 1 0

-0.00705498173140282 1 0

-0.00703011018111343 1 0

-0.00697545050223431 1 0

-0.00694857921794936 1 0

-0.00692519957633443 1 0

-0.00690686817073177 1 0

-0.00690196932174427 1 0

-0.00688055826018112 1 0

-0.00687771740974965 1 0

-0.0068460773974788 1 0

-0.00683340312417489 1 0

-0.00680699213507804 1 0

-0.00680298106886714 1 0

-0.00680152324295222 1 0

-0.00678830504408669 1 0

-0.0067841136088783 1 0

-0.0067470123571233 1 0

-0.00673770225296787 1 0

-0.00672040000444134 1 0

-0.00671438845729028 1 0

-0.006697073132894 1 0

-0.00665328479737486 1 0

-0.0066495658266108 1 0

-0.00663924012211739 1 0

-0.00661164274816541 1 0

-0.00659981267682476 1 0

-0.00659587713897603 1 0

-0.00658569908536211 1 0

-0.00656907970577283 1 0

-0.00656190076997208 1 0

-0.00655608163566642 1 0

-0.00651505929887611 1 0

-0.00650733259736321 1 0

-0.00650425928826773 1 0

-0.00649693072692766 1 0

-0.00647596913299977 1 0

-0.00646825242655058 1 0

-0.00646183502397859 1 0

-0.00645422799893614 1 0

-0.00643272538444229 1 0

-0.00640553486662987 1 0

-0.00640324150057487 1 0

-0.00638051345612593 1 0

-0.0063590163456481 1 0

-0.00634130744625105 1 0

-0.00630378094380427 1 0

-0.00630311204681206 1 0

-0.00628563222973812 1 0

-0.00625229019025916 1 0

-0.00623462089289714 1 0

-0.006211220737563 1 0

-0.00619886275292707 1 0

-0.00619555049763474 1 0

-0.006168593612058 1 0

-0.00615689805676446 1 0

-0.00615381584358973 1 0

-0.00614543952340998 1 0

-0.00613365478428745 1 0

-0.00607980256571178 1 0

-0.00606691589415635 1 0

-0.00606365296794427 1 0

-0.00604291951934006 1 0

-0.00601684225556594 1 0

-0.00601553163480465 1 0

-0.00598042293769107 1 0

-0.00597977856270035 1 0

-0.00597876452269729 1 0

-0.00594488436774579 1 0

-0.00594021653481709 1 0

-0.00592387500507258 1 0

-0.0059195865385821 1 0

-0.00584736103833476 1 0

-0.00584567014373752 1 0

-0.00584155701736059 1 0

-0.00580972258618354 1 0

-0.00579980676065283 1 0

-0.0057931169828459 1 0

-0.00579191464754661 1 0

-0.00578248581723734 1 0

-0.00578101812934568 1 0

-0.00572044902234639 1 0

-0.00568904681026238 1 0

-0.00568738935316609 1 0

-0.00560496042109185 1 0

-0.00559421872928526 1 0

-0.00558703277468125 1 0

-0.00558273123990189 1 0

-0.00557596996485544 1 0

-0.00557201768305523 1 0

-0.00554584806812181 1 0

-0.00553457823296829 1 0

-0.00550562135490065 1 0

-0.00549283099908367 1 0

-0.00544959212644013 1 0

-0.00543997285410599 1 0

-0.00543836356865975 1 0

-0.00543792586737602 1 0

-0.00542770960961412 1 0

-0.00541706034683274 1 0

-0.0054170575235423 1 0

-0.00541007868647413 1 0

-0.00538970288081442 1 0

-0.00538538767530876 1 0

-0.00538313262650794 1 0

-0.00533805499359547 1 0

-0.00533701040208609 1 0

-0.00532903773941237 1 0

-0.00530463961146482 1 0

-0.00530167152986567 1 0

-0.00529965183330434 1 0

-0.00529319322748355 1 0

-0.00528272609740661 1 0

-0.00528137121927016 1 0

-0.00527142923343935 1 0

-0.00523883417652902 1 0

-0.00523002118334581 1 0

-0.00521337843544367 1 0

-0.00519352099942522 1 0

-0.0051714132973945 1 0

-0.00516817625398613 1 0

-0.00516680841682768 1 0

-0.00515091859545445 1 0

-0.0051471506358107 1 0

-0.00514194430822708 1 0

-0.00514183241738752 1 0

-0.00513856755165813 1 0

-0.00512839103081747 1 0

-0.00511617741662456 1 0

-0.00511223302972213 1 0

-0.00506428564075133 1 0

-0.00506044919911034 1 0

-0.00504332014022716 1 0

-0.00503968474569307 1 0

-0.00503879927777686 1 0

-0.00503555896919149 1 0

-0.00503473255550372 1 0

-0.0050342289845603 1 0

-0.0050294385188815 1 0

-0.00499263089918593 1 0

-0.00498179064706555 1 0

-0.00495856811924142 1 0

-0.00494215095944872 1 0

-0.00493833897326574 1 0

-0.00492333714946131 1 0

-0.00490629678123657 1 0

-0.00489670954856161 1 0

-0.00487032179284611 1 0

-0.00486890468557004 1 0

-0.00485813291544691 1 0

-0.00485186975671484 1 0

-0.00483582940230681 1 0

-0.00483311828139238 1 0

-0.00482045179371935 1 0

-0.00481264575903178 1 0

-0.00476736186289792 1 0

-0.00476182755441914 1 0

-0.00474661581545307 1 0

-0.00474264547546187 1 0

-0.00472364514530535 1 0

-0.0046984223246536 1 0

-0.00467852815081469 1 0

-0.00467430996901025 1 0

-0.00466609225641274 1 0

-0.00463376486705291 1 0

-0.00462609800185557 1 0

-0.00458046364028682 1 0

-0.00457387429576751 1 0

-0.00456822302100252 1 0

[ reached getOption("max.print") -- omitted 7568 rows ]

**iv.**

**CALCULATION OF FP, FN, SENSITITVITY, SPECIFICITY FOR LASSO MODEL**

**FP rate =**

**FP rate =**

**SPECIFCITY =**

**SENSITIVITY =**

**MISCLASSIFICATION RATE = 0.0056**

**Question 4.**

**i.**

#pls#

> set.seed(8)

> pls.fit=plsr(nature~., data = jmtrain, scale =TRUE, validation ="CV")

> validationplot(pls.fit, val.type="MSEP")

> summary(pls.fit)

Data: X dimension: 10060 122

Y dimension: 10060 1

Fit method: kernelpls

Number of components considered: 122

VALIDATION: RMSEP

Cross-validated using 10 random segments.

(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps

CV 0.2291 0.1347 0.1251 0.1195 0.1190 0.1189

adjCV 0.2291 0.1344 0.1250 0.1194 0.1189 0.1188

6 comps 7 comps 8 comps 9 comps 10 comps 11 comps

CV 0.1190 0.1190 0.1190 0.1191 0.1192 0.1196

adjCV 0.1188 0.1188 0.1188 0.1189 0.1191 0.1194

12 comps 13 comps 14 comps 15 comps 16 comps 17 comps

CV 0.1203 0.1205 0.1201 0.1196 0.1195 0.1194

adjCV 0.1200 0.1202 0.1199 0.1194 0.1193 0.1192

18 comps 19 comps 20 comps 21 comps 22 comps 23 comps

CV 0.1193 0.1193 0.1192 0.1192 0.1192 0.1192

adjCV 0.1192 0.1191 0.1191 0.1190 0.1190 0.1190

24 comps 25 comps 26 comps 27 comps 28 comps 29 comps

CV 0.1192 0.1192 0.1192 0.1192 0.1193 0.1194

adjCV 0.1190 0.1190 0.1190 0.1190 0.1191 0.1192

30 comps 31 comps 32 comps 33 comps 34 comps 35 comps

CV 0.1194 0.1194 0.1194 0.1193 0.1193 0.1194

adjCV 0.1192 0.1192 0.1192 0.1192 0.1192 0.1192

36 comps 37 comps 38 comps 39 comps 40 comps 41 comps

CV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

adjCV 0.1192 0.1192 0.1192 0.1192 0.1193 0.1193

42 comps 43 comps 44 comps 45 comps 46 comps 47 comps

CV 0.1194 0.1195 0.1195 0.1195 0.1195 0.1195

adjCV 0.1193 0.1193 0.1193 0.1193 0.1194 0.1194

48 comps 49 comps 50 comps 51 comps 52 comps 53 comps

CV 0.1196 0.1196 0.1196 0.1195 0.1195 0.1195

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

54 comps 55 comps 56 comps 57 comps 58 comps 59 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

60 comps 61 comps 62 comps 63 comps 64 comps 65 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

66 comps 67 comps 68 comps 69 comps 70 comps 71 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

72 comps 73 comps 74 comps 75 comps 76 comps 77 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

78 comps 79 comps 80 comps 81 comps 82 comps 83 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

84 comps 85 comps 86 comps 87 comps 88 comps 89 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

90 comps 91 comps 92 comps 93 comps 94 comps 95 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

96 comps 97 comps 98 comps 99 comps 100 comps 101 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194 0.1194

102 comps 103 comps 104 comps 105 comps 106 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194

107 comps 108 comps 109 comps 110 comps 111 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194

112 comps 113 comps 114 comps 115 comps 116 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194

117 comps 118 comps 119 comps 120 comps 121 comps

CV 0.1196 0.1196 0.1196 0.1196 0.1196

adjCV 0.1194 0.1194 0.1194 0.1194 0.1194

122 comps

CV 0.1196

adjCV 0.1194

TRAINING: % variance explained

1 comps 2 comps 3 comps 4 comps 5 comps 6 comps

X 3.083 9.931 13.22 15.45 16.81 18.02

nature 66.403 70.661 73.35 73.80 73.98 74.03

7 comps 8 comps 9 comps 10 comps 11 comps 12 comps

X 18.98 20.26 21.07 21.88 22.67 23.04

nature 74.04 74.04 74.04 74.05 74.05 74.06

13 comps 14 comps 15 comps 16 comps 17 comps 18 comps

X 23.79 24.52 25.11 25.67 26.36 26.93

nature 74.07 74.07 74.07 74.07 74.08 74.08

19 comps 20 comps 21 comps 22 comps 23 comps 24 comps

X 27.69 28.37 29.07 29.71 30.20 30.83

nature 74.08 74.08 74.08 74.08 74.08 74.08

25 comps 26 comps 27 comps 28 comps 29 comps 30 comps

X 31.43 31.89 32.49 33.09 33.76 34.37

nature 74.08 74.08 74.08 74.08 74.08 74.08

31 comps 32 comps 33 comps 34 comps 35 comps 36 comps

X 34.94 35.43 35.99 36.69 37.29 37.94

nature 74.08 74.08 74.09 74.09 74.09 74.09

37 comps 38 comps 39 comps 40 comps 41 comps 42 comps

X 38.46 38.98 39.62 40.24 40.86 41.46

nature 74.09 74.09 74.09 74.09 74.09 74.09

43 comps 44 comps 45 comps 46 comps 47 comps 48 comps

X 41.89 42.28 42.90 43.57 44.13 44.81

nature 74.09 74.09 74.09 74.09 74.09 74.09

49 comps 50 comps 51 comps 52 comps 53 comps 54 comps

X 45.04 45.66 46.39 47.13 47.81 48.07

nature 74.09 74.09 74.09 74.09 74.09 74.09

55 comps 56 comps 57 comps 58 comps 59 comps 60 comps

X 48.64 49.30 49.96 50.71 51.39 51.66

nature 74.09 74.09 74.09 74.09 74.09 74.09

61 comps 62 comps 63 comps 64 comps 65 comps 66 comps

X 52.22 52.97 53.71 54.47 55.19 55.92

nature 74.09 74.09 74.09 74.09 74.09 74.09

67 comps 68 comps 69 comps 70 comps 71 comps 72 comps

X 56.64 57.41 58.20 58.96 59.74 60.49

nature 74.09 74.09 74.09 74.09 74.09 74.09

73 comps 74 comps 75 comps 76 comps 77 comps 78 comps

X 61.21 61.97 62.74 63.51 64.28 65.07

nature 74.09 74.09 74.09 74.09 74.09 74.09

79 comps 80 comps 81 comps 82 comps 83 comps 84 comps

X 65.85 66.64 67.37 68.12 68.87 69.61

nature 74.09 74.09 74.09 74.09 74.09 74.09

85 comps 86 comps 87 comps 88 comps 89 comps 90 comps

X 70.35 71.10 71.84 72.58 73.33 74.07

nature 74.09 74.09 74.09 74.09 74.09 74.09

91 comps 92 comps 93 comps 94 comps 95 comps 96 comps

X 74.81 75.56 76.30 77.04 77.78 78.53

nature 74.09 74.09 74.09 74.09 74.09 74.09

97 comps 98 comps 99 comps 100 comps 101 comps

X 79.27 80.01 80.76 81.50 82.24

nature 74.09 74.09 74.09 74.09 74.09

102 comps 103 comps 104 comps 105 comps 106 comps

X 82.99 83.73 84.47 85.22 85.96

nature 74.09 74.09 74.09 74.09 74.09

107 comps 108 comps 109 comps 110 comps 111 comps

X 86.70 87.45 88.19 88.93 89.68

nature 74.09 74.09 74.09 74.09 74.09

112 comps 113 comps 114 comps 115 comps 116 comps

X 90.42 91.16 91.91 92.65 93.39

nature 74.09 74.09 74.09 74.09 74.09

117 comps 118 comps 119 comps 120 comps 121 comps

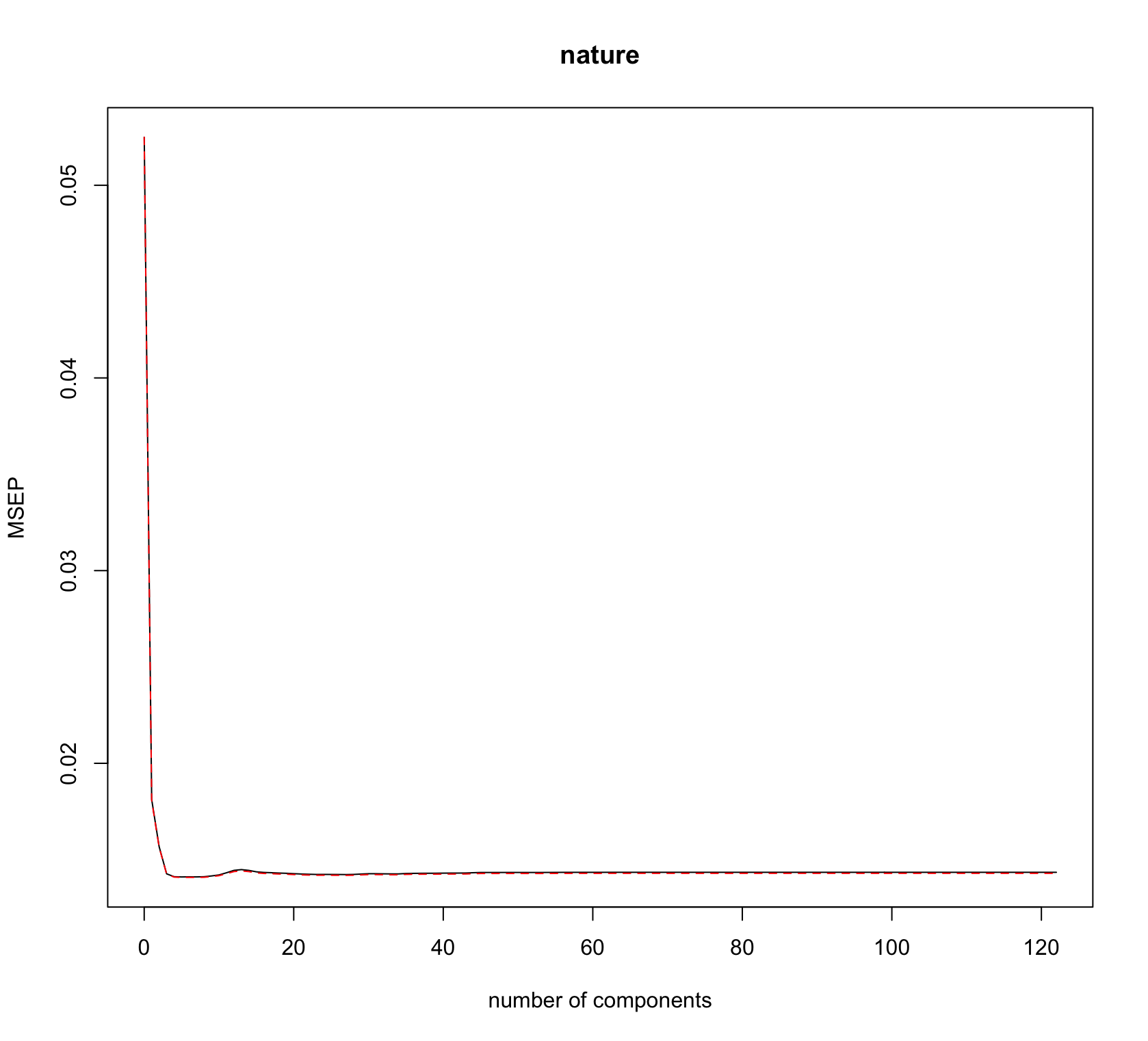
X 94.14 94.88 95.62 96.36 97.11

nature 74.09 74.09 74.09 74.09 74.09

122 comps

X 97.85

nature 74.09

ii. 

iii.

> mean((fit.pls.pred-ytest)^2)

[1] 0.01742148

(b)

i.

> set.seed(8)

> pcr.fit=pcr(nature~.,data=jmtrain, scale=TRUE, validation = "CV")

> validationplot(pcr.fit, val.type = "MSEP")

> summary(pcr.fit)

Data: X dimension: 10060 122

Y dimension: 10060 1

Fit method: svdpc

Number of components considered: 122

VALIDATION: RMSEP

Cross-validated using 10 random segments.

(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps

CV 0.2291 0.2288 0.2075 0.1660 0.1525 0.1412

adjCV 0.2291 0.2288 0.2069 0.1659 0.1516 0.1410

6 comps 7 comps 8 comps 9 comps 10 comps 11 comps

CV 0.1285 0.1235 0.1231 0.1228 0.1225 0.1223

adjCV 0.1266 0.1238 0.1229 0.1226 0.1222 0.1221

12 comps 13 comps 14 comps 15 comps 16 comps 17 comps

CV 0.1222 0.1222 0.1222 0.1222 0.1221 0.1221

adjCV 0.1221 0.1221 0.1221 0.1220 0.1220 0.1219

18 comps 19 comps 20 comps 21 comps 22 comps 23 comps

CV 0.1220 0.1220 0.1221 0.122 0.122 0.1219

adjCV 0.1219 0.1219 0.1220 0.122 0.122 0.1219

24 comps 25 comps 26 comps 27 comps 28 comps 29 comps

CV 0.1218 0.1216 0.1215 0.1214 0.1214 0.1214

adjCV 0.1217 0.1217 0.1212 0.1212 0.1212 0.1212

30 comps 31 comps 32 comps 33 comps 34 comps 35 comps

CV 0.1214 0.1213 0.1213 0.1212 0.1213 0.1213

adjCV 0.1212 0.1212 0.1211 0.1211 0.1212 0.1212

36 comps 37 comps 38 comps 39 comps 40 comps 41 comps

CV 0.1213 0.1213 0.1213 0.1213 0.1212 0.1212

adjCV 0.1211 0.1211 0.1211 0.1212 0.1211 0.1211

42 comps 43 comps 44 comps 45 comps 46 comps 47 comps

CV 0.1213 0.1212 0.1212 0.1212 0.1212 0.1212

adjCV 0.1212 0.1211 0.1211 0.1211 0.1211 0.1211

48 comps 49 comps 50 comps 51 comps 52 comps 53 comps

CV 0.1212 0.1212 0.1212 0.1212 0.1211 0.1211

adjCV 0.1211 0.1211 0.1211 0.1211 0.1211 0.1210

54 comps 55 comps 56 comps 57 comps 58 comps 59 comps

CV 0.1211 0.1211 0.1211 0.1211 0.121 0.121

adjCV 0.1211 0.1211 0.1210 0.1210 0.121 0.121

60 comps 61 comps 62 comps 63 comps 64 comps 65 comps

CV 0.121 0.1209 0.1210 0.1210 0.1210 0.1209

adjCV 0.121 0.1209 0.1209 0.1209 0.1209 0.1208

66 comps 67 comps 68 comps 69 comps 70 comps 71 comps

CV 0.1210 0.1210 0.1209 0.1209 0.1210 0.1211

adjCV 0.1209 0.1209 0.1208 0.1209 0.1209 0.1209

72 comps 73 comps 74 comps 75 comps 76 comps 77 comps

CV 0.1211 0.1211 0.1211 0.1211 0.1212 0.1212

adjCV 0.1210 0.1210 0.1210 0.1211 0.1212 0.1212

78 comps 79 comps 80 comps 81 comps 82 comps 83 comps

CV 0.1211 0.1209 0.1209 0.1207 0.1206 0.1203

adjCV 0.1209 0.1208 0.1208 0.1207 0.1207 0.1203

84 comps 85 comps 86 comps 87 comps 88 comps 89 comps

CV 0.1203 0.1202 0.1203 0.1202 0.1201 0.1200

adjCV 0.1202 0.1200 0.1202 0.1201 0.1201 0.1199

90 comps 91 comps 92 comps 93 comps 94 comps 95 comps

CV 0.1200 0.1200 0.1199 0.1200 0.1200 0.1200

adjCV 0.1199 0.1199 0.1199 0.1199 0.1199 0.1199

96 comps 97 comps 98 comps 99 comps 100 comps 101 comps

CV 0.1200 0.1200 0.1199 0.1201 0.1202 0.1201

adjCV 0.1199 0.1199 0.1198 0.1200 0.1202 0.1200

102 comps 103 comps 104 comps 105 comps 106 comps

CV 0.1197 0.1198 0.1196 0.1190 0.1190

adjCV 0.1198 0.1204 0.1194 0.1189 0.1189

107 comps 108 comps 109 comps 110 comps 111 comps

CV 0.1190 0.1190 0.1191 0.1189 0.1189

adjCV 0.1189 0.1189 0.1190 0.1188 0.1188

112 comps 113 comps 114 comps 115 comps 116 comps

CV 0.1190 0.1192 0.1194 0.1193 0.1193

adjCV 0.1189 0.1191 0.1192 0.1192 0.1191

117 comps 118 comps 119 comps 120 comps 121 comps

CV 0.1193 0.1192 0.1192 0.1193 0.1193

adjCV 0.1191 0.1191 0.1190 0.1191 0.1191

122 comps

CV 0.1196

adjCV 0.1194

TRAINING: % variance explained

1 comps 2 comps 3 comps 4 comps 5 comps 6 comps

X 8.2952 11.64 14.31 16.21 17.95 19.34

nature 0.3315 19.95 47.51 56.13 62.12 70.85

7 comps 8 comps 9 comps 10 comps 11 comps 12 comps

X 20.62 21.86 23.04 24.18 25.31 26.42

nature 70.86 71.33 71.55 71.72 71.77 71.77

13 comps 14 comps 15 comps 16 comps 17 comps 18 comps

X 27.49 28.52 29.49 30.46 31.42 32.37

nature 71.78 71.83 71.87 71.88 71.91 71.93

19 comps 20 comps 21 comps 22 comps 23 comps 24 comps

X 33.31 34.25 35.19 36.11 37.02 37.93

nature 71.94 71.94 71.95 71.95 71.96 72.05

25 comps 26 comps 27 comps 28 comps 29 comps 30 comps

X 38.84 39.74 40.63 41.53 42.41 43.30

nature 72.05 72.27 72.27 72.27 72.28 72.28

31 comps 32 comps 33 comps 34 comps 35 comps 36 comps

X 44.18 45.05 45.92 46.80 47.67 48.53

nature 72.29 72.30 72.30 72.31 72.31 72.35

37 comps 38 comps 39 comps 40 comps 41 comps 42 comps

X 49.39 50.25 51.10 51.95 52.79 53.63

nature 72.35 72.36 72.36 72.37 72.37 72.38

43 comps 44 comps 45 comps 46 comps 47 comps 48 comps

X 54.48 55.31 56.15 56.98 57.80 58.62

nature 72.40 72.42 72.42 72.46 72.47 72.48

49 comps 50 comps 51 comps 52 comps 53 comps 54 comps

X 59.44 60.25 61.07 61.87 62.68 63.48

nature 72.49 72.49 72.49 72.50 72.50 72.50

55 comps 56 comps 57 comps 58 comps 59 comps 60 comps

X 64.28 65.08 65.87 66.65 67.44 68.22

nature 72.50 72.52 72.52 72.54 72.55 72.56

61 comps 62 comps 63 comps 64 comps 65 comps 66 comps

X 69.00 69.78 70.55 71.32 72.08 72.84

nature 72.57 72.60 72.60 72.61 72.64 72.65

67 comps 68 comps 69 comps 70 comps 71 comps 72 comps

X 73.60 74.36 75.11 75.86 76.60 77.35

nature 72.66 72.66 72.66 72.71 72.72 72.72

73 comps 74 comps 75 comps 76 comps 77 comps 78 comps

X 78.09 78.82 79.55 80.28 81.00 81.72

nature 72.73 72.75 72.75 72.75 72.76 72.85

79 comps 80 comps 81 comps 82 comps 83 comps 84 comps

X 82.43 83.14 83.85 84.55 85.24 85.93

nature 72.85 72.86 72.86 72.86 72.99 73.06

85 comps 86 comps 87 comps 88 comps 89 comps 90 comps

X 86.61 87.29 87.95 88.60 89.26 89.9

nature 73.13 73.13 73.14 73.14 73.19 73.2

91 comps 92 comps 93 comps 94 comps 95 comps 96 comps

X 90.53 91.16 91.77 92.37 92.96 93.52

nature 73.20 73.21 73.21 73.21 73.21 73.22

97 comps 98 comps 99 comps 100 comps 101 comps

X 94.08 94.62 95.15 95.68 96.17

nature 73.22 73.24 73.25 73.25 73.36

102 comps 103 comps 104 comps 105 comps 106 comps

X 96.63 97.06 97.50 97.88 98.24

nature 73.37 73.37 73.96 73.96 73.97

107 comps 108 comps 109 comps 110 comps 111 comps

X 98.57 98.89 99.18 99.42 99.66

nature 73.97 73.97 73.97 74.02 74.04

112 comps 113 comps 114 comps 115 comps 116 comps

X 99.84 99.92 99.97 99.98 99.99

nature 74.04 74.04 74.04 74.04 74.05

117 comps 118 comps 119 comps 120 comps 121 comps

X 99.99 100.00 100.00 100.00 100.00

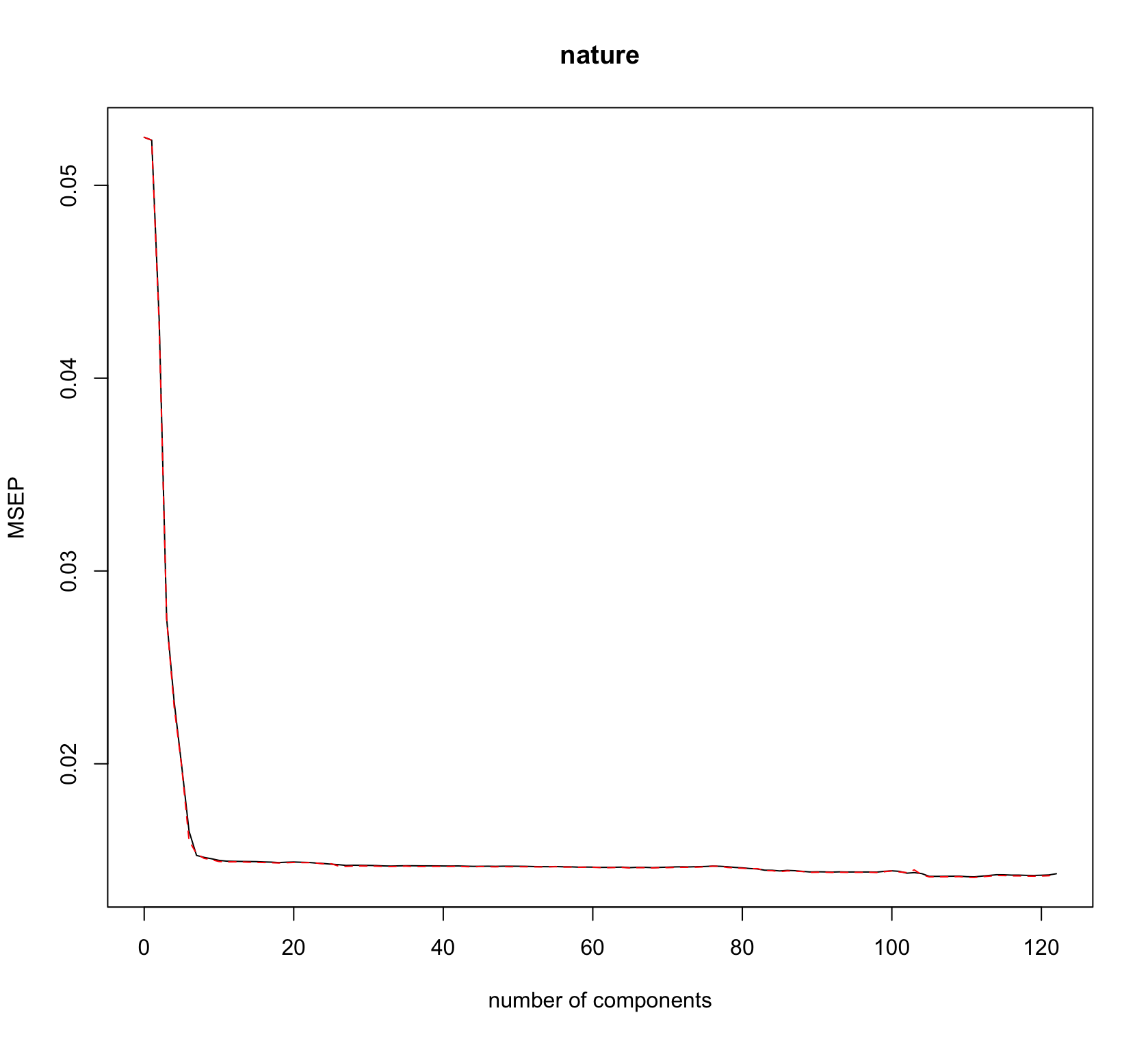
nature 74.07 74.07 74.08 74.08 74.09

122 comps

X 100.00

nature 74.09

ii.



iii.

> pred.pcr = predict(pcr.fit, xtest, ncomp=3)

> mean((pred.pcr-ytest)^2)

[1] 0.03123046

**THE DIFFERENCE BETWEEN PLS AND PCR**

**In the PCR, the main concentration is to identify linear combination of variables that explain or depicts the best representation of the model. The response variables which is not used to represent the PCR is omitted. On the other hand, PCR has no level of assurance that the best explanatory variable will be predicted in the predicting the response of our model**

**Question 5.**

**(a)**

**i.**

> #lda Model#

> lda.fit = lda(nature~., data=jmtrain)

> lda.fit

Call:

lda(nature ~ ., data = jmtrain)

Prior probabilities of groups:

0 1

0.9444334 0.0555666

Group means:

x1 x2 x3 x4 x5 x6

0 11.784549 11.781286 0.09062204 1.784023 114216.00 3077.7614

1 6.822898 6.838998 0.08407871 1.688730 40505.37 282.5886

x7 x8 x9 x10 x11 x12

0 110909.53 7967.842 530.84075 289.18103 217.95685 509.26302

1 39615.07 2651.795 89.13238 34.47406 39.33631 74.39535

x13 x14 x15 x16 x17 x18 x19

0 1.477213 0.09125355 4058.524 49168.822 22.996842 421.7558 352.3510

1 1.456172 0.08228980 1530.592 3451.558 4.323792 178.5338 140.5515

x42 x43 x44 x45 x46 x47

0 0.08893801 0.6618251 18.62330 79.48332 0.09272708 64155.31

1 0.09481216 0.6440072 24.59213 79.39356 0.08765653 23259.55

x48 x49 x50 x51 x52 x53

0 0.09167456 0.08851700 0.09041154 0.09251658 0.09072729 0.09062204

1 0.07692308 0.09302326 0.08765653 0.10375671 0.09302326 0.09838998

x54 x55 x56 x57 x58 x59

0 0.09272708 0.08914851 0.4806862 0.09041154 0.09430586 0.09346385

1 0.07155635 0.09838998 0.4096601 0.08228980 0.10375671 0.09660107

x60 x61 x62 x63 x64 x65

0 0.09177981 0.08788549 18.62372 0.7288706 1.507105 0.09262183

1 0.09481216 0.09481216 24.61360 1.2397138 1.676208 0.11091234

x66 x67 x68 x69 x70 x71

0 0.09430586 0.1183033 0.08725397 0.2225029 0.2774445 1.0123145

1 0.08944544 0.1592129 0.10912343 0.7513417 0.1842576 0.5384615

x72 x73 x74 x75 x76 x77

0 0.4009052 0.11661930 0.08767498 0.2870224 0.1899800 0.10272603

1 0.1162791 0.07871199 0.07155635 0.1109123 0.1162791 0.08944544

x78 x79 x80 x81 x82 x83

0 0.7056099 0.10146300 0.1658773 0.11567203 0.09820019 0.3322808

1 0.3989267 0.08407871 0.1055456 0.07155635 0.08765653 0.1806798

x84 x85 x86 x87 x88 x89

0 0.1699821 0.5224713 2.399011 0.2540785 2.043785 0.1336701

1 0.1288014 0.1413238 1.325581 0.1073345 3.524150 0.1109123

x90 x91 x92 x93 x94 x95

0 0.2286075 0.13072308 0.3242817 0.09935796 0.8407536 0.10209452

1 0.1377460 0.07334526 0.1001789 0.10912343 0.1270125 0.09838998

x96 x97 x98 x99 x100 x101

0 0.33522787 0.1832439 0.13609094 0.09883170 0.09514788 1.1994527

1 0.07334526 0.1216458 0.09481216 0.08586762 0.10733453 0.7423971

x102 x103 x104 x105 x106 x107

0 0.8301231 2.3204926 0.8505420 0.1033575 0.11682981 0.32091359

1 0.1395349 0.8121646 0.3220036 0.0822898 0.09302326 0.09838998

x108 x109 x110 x111 x112 x113

0 0.09230607 5.934323 0.1849279 0.9909483 0.4210083 8.900221

1 0.10017889 1.776386 0.2128801 0.3452594 0.1949911 2.894454

x114 x115 x116 x117 x118 x119

0 0.5280497 1.8468582 0.7300284 0.2193453 3.0591517 33.03673

1 0.2093023 0.7871199 0.3989267 0.2307692 0.6100179 19.05188

x120 x121 x122 x123 x124 x201

0 1.4744764 46.17524 9.988212 11.9043259 0.09683191 0.7042417

1 0.6529517 19.73166 1.744186 0.2450805 0.08228980 0.9141324

x202 x203 x204 x205 x206 x207

0 0.9142196 0.003157562 0.5089991 0.4979476 0.09714767 0.001368277

1 0.6905188 1.026833631 0.6064401 0.6493739 0.00000000 0.699463327

x208 x209 x210 x211 x212 x213

0 0.7039259 0.001368277 0.8981160 0.1996632 0.4973161 0.09956847

1 1.0125224 0.472271914 0.4991055 0.3899821 0.8354204 0.56887299

x214 x215 x216 x217 x218 x219

0 0.1009367 0.4991054 0.8072834 0.7930744 0.002315546 9.076023

1 0.2164580 0.3720930 0.5957066 0.7584973 0.599284436 10.791373

x220

0 0.5013380

1 0.4032638

Coefficients of linear discriminants:

LD1

x1 1.683431e-02

x2 -1.421470e-02

x3 4.549963e-02

x4 3.087136e-02

x5 1.687473e-06

x6 -6.112538e-07

x7 -1.569579e-06

x8 -5.633178e-05

x9 -4.438632e-05

x10 2.098418e-04

x11 1.734212e-04

x12 -2.559925e-04

x13 6.441517e-04

x14 -8.609871e-02

x15 -8.428549e-06

x16 2.026017e-07

x17 5.863821e-04

x18 1.044382e-03

x19 -1.315723e-03

x42 -4.782773e-03

x43 1.708992e-03

x44 3.916923e-02

x45 -5.631678e-03

x46 -1.787480e-02

x47 -3.837001e-07

x48 -1.145559e-02

x49 3.298732e-03

x50 6.700045e-02

x51 -1.080842e-02

x52 1.663581e-02

x53 -6.698552e-02

x54 -3.642499e-02

x55 2.916943e-02

x56 -1.081016e-01

x57 3.914850e-02

x58 -3.054887e-02

x59 -1.046245e-02

x60 2.943185e-02

x61 2.801806e-03

x62 1.590349e-02

x63 1.751668e-01

x64 4.566564e-02

x65 1.481086e-02

x66 -8.377972e-02

x67 2.850119e-02

x68 2.258634e-02

x69 7.991168e-02

x70 -1.031690e-03

x71 2.037423e-04

x72 1.528675e-03

x73 -5.114800e-02

x74 -1.774733e-02

x75 -6.824387e-03

x76 -3.170940e-03

x77 -2.960175e-02

x78 -2.311899e-02

x79 2.948589e-03

x80 -1.228574e-02

x81 -1.186531e-02

x82 3.904836e-03

x83 -2.136247e-02

x84 2.663861e-03

x85 2.086970e-04

x86 -1.208927e-04

x87 -5.580547e-03

x88 1.125405e-03

x89 -2.703670e-02

x90 6.872185e-03

x91 -1.877423e-02

x92 -3.537073e-03

x93 -1.039875e-03

x94 -1.898759e-04

x95 3.050209e-02

x96 -5.286116e-04

x97 -2.532924e-04

x98 -1.722117e-02

x99 -8.476485e-03

x100 1.828859e-02

x101 1.151089e-04

x102 -8.891849e-04

x103 -1.789832e-04

x104 -1.159937e-03

x105 -6.522746e-03

x106 -8.556694e-03

x107 -1.161600e-03

x108 3.417836e-02

x109 -2.076979e-03

x110 1.418168e-02

x111 -5.349311e-03

x112 -4.472167e-03

x113 -2.561476e-04

x114 -2.564085e-03

x115 1.052370e-03

x116 -4.116467e-04

x117 3.166341e-03

x118 3.300782e-04

x119 2.602912e-04

x120 -6.002929e-04

x121 -1.754065e-04

x122 -3.246281e-04

x123 1.051530e-03

x124 -6.496403e-02

x201 4.429057e-02

x202 -2.857588e-02

x203 2.250026e+00

x204 2.999498e-02

x205 4.972272e-02

x206 -1.287294e-01

x207 2.332405e+00

x208 6.657051e-02

x209 2.124878e+00

x210 -5.600196e-02

x211 1.436957e-01

x212 7.062137e-02

x213 4.217610e-01

x214 2.029619e-01

x215 -1.230770e-02

x216 -2.415148e-02

x217 -5.153908e-03

x218 1.978661e+00

x219 1.259842e-02

x220 -1.636497e-01

ii.

> pred.fit.lda = predict(lda.fit, jmtest)

>

> mean(ytest!=pred.fit.lda$class)

[1] 0.01536936

iii.

> table(pred.fit.lda$class, ytest)

ytest

0 1

0 7482 105

1. 19 462

**MISCLASSIFICATION RATE = 0.0153**

**SUMMARY OF RESULTS**

**FP Rates**

**In this particular scenario, we conclude that percentage of correct predictions on the training data is (54+557)/1089 which is equal to 56.1065%. This can be also be interpreted as 43.8935% is the training error rate, which is often overly optimistic. The overall accuracy of the model is 56,11%. The sensitivity is 56.43% which explains that we are able to perform better than** the baseline

**FP Rates**

**In this particular scenario, we may conclude that the percentage of correct predictions on the test data is 62.5%.In other words, using the most significant variable, in our logistic regression model we find the overall accuracy of the model to be (9+56)/104 which is equal to 62.5%**

**Question 6.**

**i.**

> #Logistic Model#

> logistic.model=glm(nature~., data=jmtrain, family=binomial)

Warning messages:

1: glm.fit: algorithm did not converge

2: glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(logistic.model)

Call:

glm(formula = nature ~ ., family = binomial, data = jmtrain)

Deviance Residuals:

Min 1Q Median 3Q Max

-8.49 0.00 0.00 0.00 8.49

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -7.642e+14 2.393e+07 -31926434 <2e-16 \*\*\*

x1 -1.101e+13 7.733e+05 -14238717 <2e-16 \*\*\*

x2 1.202e+13 7.623e+05 15765297 <2e-16 \*\*\*

x3 -1.876e+13 2.352e+06 -7976183 <2e-16 \*\*\*

x4 -6.430e+12 9.864e+05 -6518406 <2e-16 \*\*\*

x5 2.535e+09 5.801e+01 43705973 <2e-16 \*\*\*

x6 -8.034e+09 1.074e+02 -74786698 <2e-16 \*\*\*

x7 -2.078e+09 6.156e+01 -33750242 <2e-16 \*\*\*

x8 5.083e+09 3.178e+02 15993995 <2e-16 \*\*\*

x9 2.242e+11 7.154e+03 31343482 <2e-16 \*\*\*

x10 5.278e+11 9.057e+03 58282033 <2e-16 \*\*\*

x11 -1.391e+12 1.285e+04 -108291875 <2e-16 \*\*\*

x12 -1.819e+10 7.473e+03 -2434236 <2e-16 \*\*\*

x13 -2.795e+13 1.313e+06 -21282411 <2e-16 \*\*\*

x14 -4.688e+13 2.321e+06 -20199549 <2e-16 \*\*\*

x15 -5.678e+09 2.100e+02 -27033290 <2e-16 \*\*\*

x16 4.242e+08 8.406e+00 50459063 <2e-16 \*\*\*

x17 -5.069e+11 2.878e+04 -17613851 <2e-16 \*\*\*

x18 9.289e+12 2.040e+04 455347686 <2e-16 \*\*\*

x19 -1.290e+13 2.719e+04 -474557596 <2e-16 \*\*\*

x42 -2.144e+13 2.361e+06 -9080236 <2e-16 \*\*\*

x43 4.224e+13 1.421e+06 29720356 <2e-16 \*\*\*

x44 2.073e+12 8.063e+05 2570913 <2e-16 \*\*\*

x45 -6.030e+12 2.935e+05 -20545780 <2e-16 \*\*\*

x46 -3.115e+13 2.325e+06 -13399484 <2e-16 \*\*\*

x47 -3.109e+08 1.245e+01 -24975508 <2e-16 \*\*\*

x48 1.434e+13 2.343e+06 6120040 <2e-16 \*\*\*

x49 2.718e+13 2.365e+06 11493641 <2e-16 \*\*\*

x50 1.053e+14 2.350e+06 44831891 <2e-16 \*\*\*

x51 -2.885e+13 2.316e+06 -12455842 <2e-16 \*\*\*

x52 -2.344e+13 2.343e+06 -10004612 <2e-16 \*\*\*

x53 -1.017e+14 2.339e+06 -43459970 <2e-16 \*\*\*

x54 -5.903e+13 2.335e+06 -25274393 <2e-16 \*\*\*

x55 -7.603e+13 2.355e+06 -32286556 <2e-16 \*\*\*

x56 -1.439e+14 1.343e+06 -107139574 <2e-16 \*\*\*

x57 -1.529e+13 2.353e+06 -6496214 <2e-16 \*\*\*

x58 -7.626e+13 2.296e+06 -33216464 <2e-16 \*\*\*

x59 -1.308e+14 2.316e+06 -56493247 <2e-16 \*\*\*

x60 -3.700e+13 2.328e+06 -15890998 <2e-16 \*\*\*

x61 -4.991e+13 2.376e+06 -21009156 <2e-16 \*\*\*

x62 3.213e+13 8.054e+05 39893697 <2e-16 \*\*\*

x63 1.199e+14 1.213e+06 98818749 <2e-16 \*\*\*

x64 3.375e+13 1.288e+06 26195477 <2e-16 \*\*\*

x65 -4.581e+12 2.308e+06 -1984687 <2e-16 \*\*\*

x66 -5.112e+13 2.306e+06 -22165992 <2e-16 \*\*\*

x67 -4.709e+13 2.029e+06 -23210043 <2e-16 \*\*\*

x68 1.073e+14 2.371e+06 45272292 <2e-16 \*\*\*

x69 2.511e+13 7.875e+05 31885115 <2e-16 \*\*\*

x70 -2.293e+12 2.374e+05 -9659645 <2e-16 \*\*\*

x71 4.350e+11 2.707e+04 16074154 <2e-16 \*\*\*

x72 1.800e+12 2.529e+05 7116610 <2e-16 \*\*\*

x73 -8.469e+13 1.558e+06 -54350446 <2e-16 \*\*\*

x74 -5.397e+12 2.390e+06 -2258272 <2e-16 \*\*\*

x75 4.930e+12 3.943e+05 12503301 <2e-16 \*\*\*

x76 -2.113e+13 4.151e+05 -50898473 <2e-16 \*\*\*

x77 -7.142e+13 1.301e+06 -54910730 <2e-16 \*\*\*

x78 -8.912e+10 4.612e+05 -193235 <2e-16 \*\*\*

x79 -1.978e+13 2.071e+06 -9553943 <2e-16 \*\*\*

x80 -4.259e+13 1.124e+06 -37885556 <2e-16 \*\*\*

x81 -2.194e+13 1.069e+06 -20522848 <2e-16 \*\*\*

x82 3.781e+13 1.573e+06 24039190 <2e-16 \*\*\*

x83 -6.073e+13 6.492e+05 -93544827 <2e-16 \*\*\*

x84 9.100e+12 4.630e+05 19655748 <2e-16 \*\*\*

x85 -5.294e+12 1.422e+05 -37237349 <2e-16 \*\*\*

x86 -4.759e+11 4.664e+04 -10202743 <2e-16 \*\*\*

x87 -2.302e+13 2.692e+05 -85537933 <2e-16 \*\*\*

x88 1.203e+12 3.848e+04 31262251 <2e-16 \*\*\*

x89 2.645e+13 1.276e+06 20725086 <2e-16 \*\*\*

x90 -1.983e+12 5.230e+05 -3792252 <2e-16 \*\*\*

x91 -4.764e+13 9.669e+05 -49269888 <2e-16 \*\*\*

x92 -2.169e+13 5.869e+05 -36964068 <2e-16 \*\*\*

x93 2.705e+13 1.609e+06 16810731 <2e-16 \*\*\*

x94 -1.411e+12 3.105e+04 -45437118 <2e-16 \*\*\*

x95 6.810e+13 2.103e+06 32390757 <2e-16 \*\*\*

x96 -1.377e+13 1.394e+05 -98761988 <2e-16 \*\*\*

x97 -1.865e+13 4.854e+05 -38424256 <2e-16 \*\*\*

x98 1.009e+13 1.173e+06 8600679 <2e-16 \*\*\*

x99 1.956e+12 1.779e+06 1099327 <2e-16 \*\*\*

x100 -4.942e+13 2.252e+06 -21949780 <2e-16 \*\*\*

x101 -1.404e+12 5.156e+04 -27229781 <2e-16 \*\*\*

x102 -1.361e+13 1.278e+05 -106501667 <2e-16 \*\*\*

x103 7.630e+11 5.023e+04 15188705 <2e-16 \*\*\*

x104 -3.563e+12 1.090e+05 -32694759 <2e-16 \*\*\*

x105 -7.406e+13 1.698e+06 -43615369 <2e-16 \*\*\*

x106 -4.785e+13 8.524e+05 -56127296 <2e-16 \*\*\*

x107 -2.936e+12 1.611e+05 -18228409 <2e-16 \*\*\*

x108 1.320e+14 2.230e+06 59220782 <2e-16 \*\*\*

x109 -4.986e+12 8.896e+04 -56046589 <2e-16 \*\*\*

x110 2.646e+12 8.288e+05 3192913 <2e-16 \*\*\*

x111 -1.537e+12 2.561e+05 -6000591 <2e-16 \*\*\*

x112 -1.175e+13 2.883e+05 -40773048 <2e-16 \*\*\*

x113 -7.446e+11 5.951e+04 -12512086 <2e-16 \*\*\*

x114 -2.215e+13 4.011e+05 -55236514 <2e-16 \*\*\*

x115 -4.590e+11 1.074e+05 -4273425 <2e-16 \*\*\*

x116 7.329e+11 2.880e+05 2544878 <2e-16 \*\*\*

x117 -7.619e+12 5.007e+05 -15217357 <2e-16 \*\*\*

x118 -2.144e+12 6.149e+04 -34870145 <2e-16 \*\*\*

x119 3.096e+11 1.345e+04 23019242 <2e-16 \*\*\*

x120 -2.112e+12 8.983e+04 -23509355 <2e-16 \*\*\*

x121 -2.383e+11 8.831e+03 -26988847 <2e-16 \*\*\*

x122 -2.191e+12 3.287e+04 -66657937 <2e-16 \*\*\*

x123 -2.294e+12 3.872e+04 -59249063 <2e-16 \*\*\*

x124 -1.221e+14 2.285e+06 -53418940 <2e-16 \*\*\*

x201 4.611e+13 8.035e+05 57389111 <2e-16 \*\*\*

x202 -4.560e+13 7.129e+05 -63966583 <2e-16 \*\*\*

x203 5.055e+14 2.432e+06 207882451 <2e-16 \*\*\*

x204 -1.760e+13 9.463e+05 -18601176 <2e-16 \*\*\*

x205 8.683e+11 9.539e+05 910287 <2e-16 \*\*\*

x206 -1.048e+15 2.251e+06 -465478515 <2e-16 \*\*\*

x207 5.147e+14 3.051e+06 168729426 <2e-16 \*\*\*

x208 8.462e+13 7.947e+05 106484778 <2e-16 \*\*\*

x209 4.238e+14 4.034e+06 105038589 <2e-16 \*\*\*

x210 -7.452e+13 7.239e+05 -102949304 <2e-16 \*\*\*

x211 8.074e+13 1.461e+06 55266181 <2e-16 \*\*\*

x212 7.456e+13 9.325e+05 79949185 <2e-16 \*\*\*

x213 1.859e+14 1.903e+06 97678521 <2e-16 \*\*\*

x214 2.004e+14 2.050e+06 97766370 <2e-16 \*\*\*

x215 -1.582e+13 9.567e+05 -16531368 <2e-16 \*\*\*

x216 -3.304e+12 7.536e+05 -4383821 <2e-16 \*\*\*

x217 -1.470e+13 7.594e+05 -19361237 <2e-16 \*\*\*

x218 4.228e+14 3.280e+06 128884263 <2e-16 \*\*\*

x219 1.883e+13 1.564e+05 120378223 <2e-16 \*\*\*

x220 -1.894e+14 2.384e+06 -79442274 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4317.6 on 10059 degrees of freedom

Residual deviance: 3892.7 on 9937 degrees of freedom

AIC: 4138.7

Number of Fisher Scoring iterations: 25

**iii.**

> log.prob = predict(logistic.model, newdata = jmtest)

> log.matrix = rep("0", nrow(jmtest))

> log.matrix[log.prob > 0.5] = "1"

> table(log.matrix, jmtest$nature)

log.matrix 0 1

0 7480 38

1 21 529

**iv.**

**CALCULATION OF FP, FN, SENSITITVITY, SPECIFICITY FOR LOGISTIC MODEL**

**FP rate =**

**FN rate =**

**SPECIFCITY = %**

**SENSITIVITY =**

**MISCLASSIFICATION RATE = 0.0073%**

**Question 7.**

**i.**

> #KNN Model#

> set.seed(1)

> pred.knn = knn(xtrain, xtest, ytrain, k=5)

> mean(ytest!=pred.knn)

[1] 0.06680714

**ii.**

> table(pred.knn,ytest)

ytest

pred.knn 0 1

0 7414 452

1 87 115

**iv.**

**CALCULATION OF FP, FN, SENSITITVITY, SPECIFICITY FOR KNN MODEL**

**FN rate =**

**FP rate =**

**SPECIFCITY = %**

**SENSITIVITY =**

**MISCLASSIFICATION RATE = 0.06680**

**EXPLANATORY SUMMARY ON RATES (FP, FN, SENSITIVITY AND SPECIFICITY)**

**From the above rates in our selected models. The ideal model show in our analysis shows Low False Negative in all of our models and same for all low a False Positive(FP). Although all the predictors above are significant in our model. Looking at FN there is drastic drop in value but this doesn’t affect any comparison made earlier. The KNN and logistic model are highly sensitive comparatively to specificity in our model hence, our conclusion will be geared towards Logistic regression since it has least FN with respect to tracking malicious activities on the website.**

**Question 11.**

> #AUC#

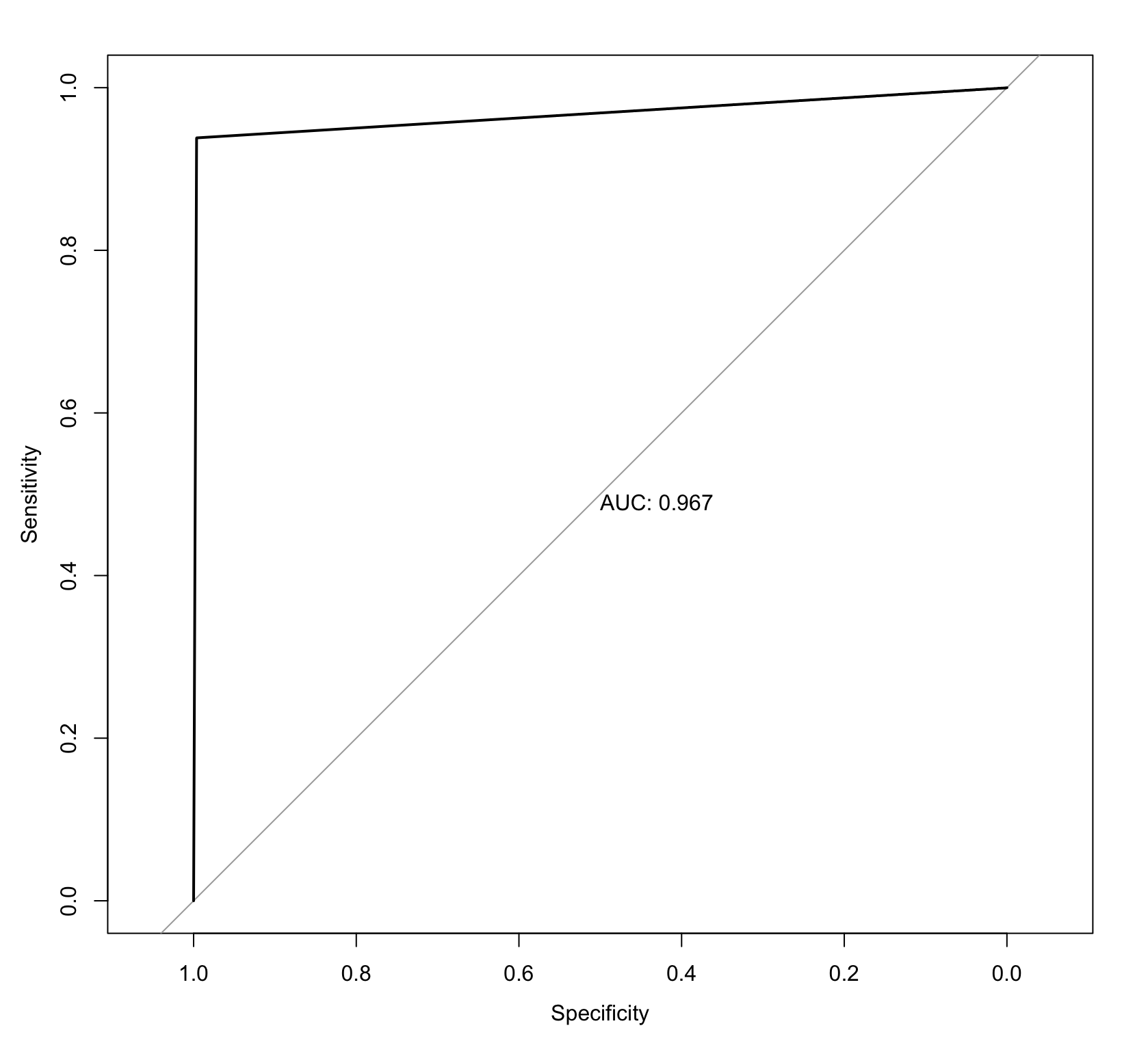
> lin.model= glm(nature~., family = "binomial", data=jmtest)

Warning message:

glm.fit: fitted probabilities numerically 0 or 1 occurred

> probs = predict(lin.model,type = c("response"))

> plot(roc(jmtest$nature,probs),print.auc=TRUE)



**Question 12.**

**FP rate =**

**FN rate =**

**SPECIFCITY = %**

**SENSITIVITY =**

**MISCLASSIFICATION RATE = 0.01680**

If we decide to control FN rate to be less than 0.01, we proceed by reducing the number of False Negative(FN) values that our model predicts. This signifies that if we reduce the constraint of threshold from 50 percent. This will decrease the number of observation that will pass malicious detection.

We will use PLS as an example.