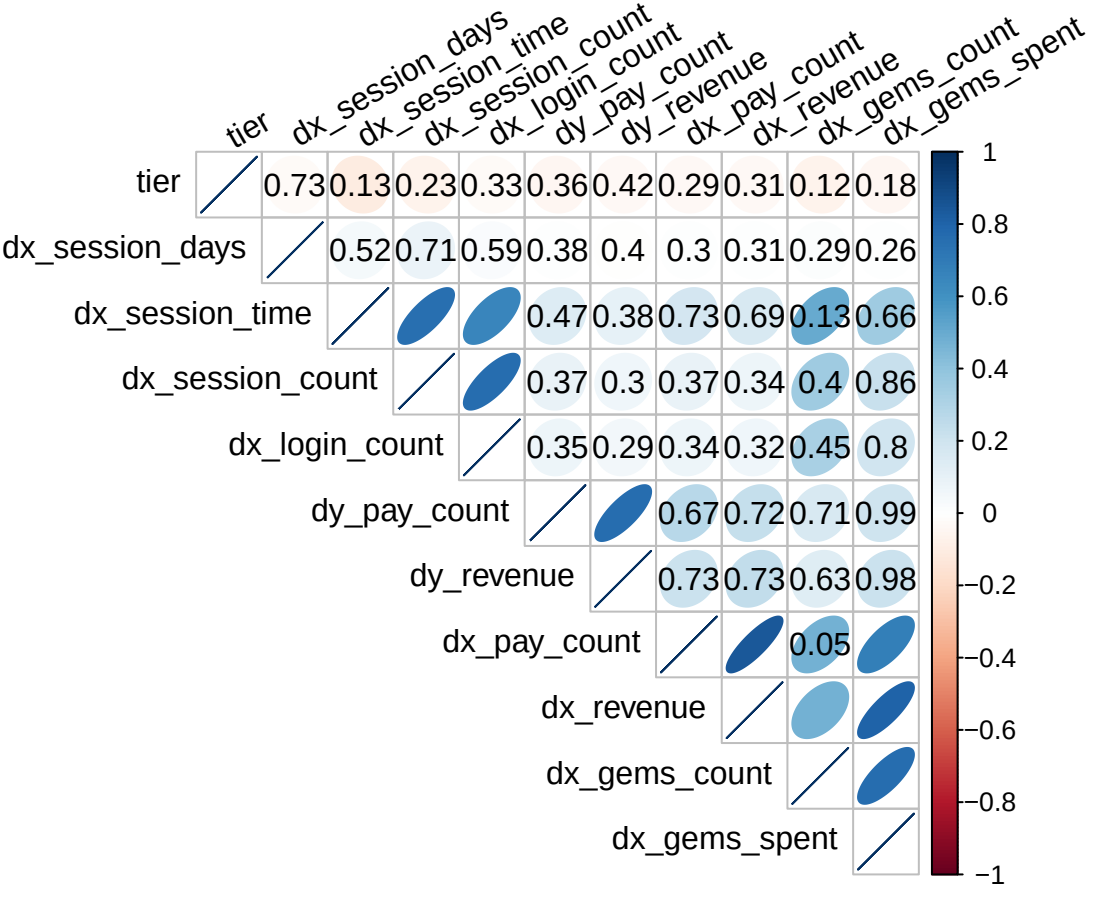
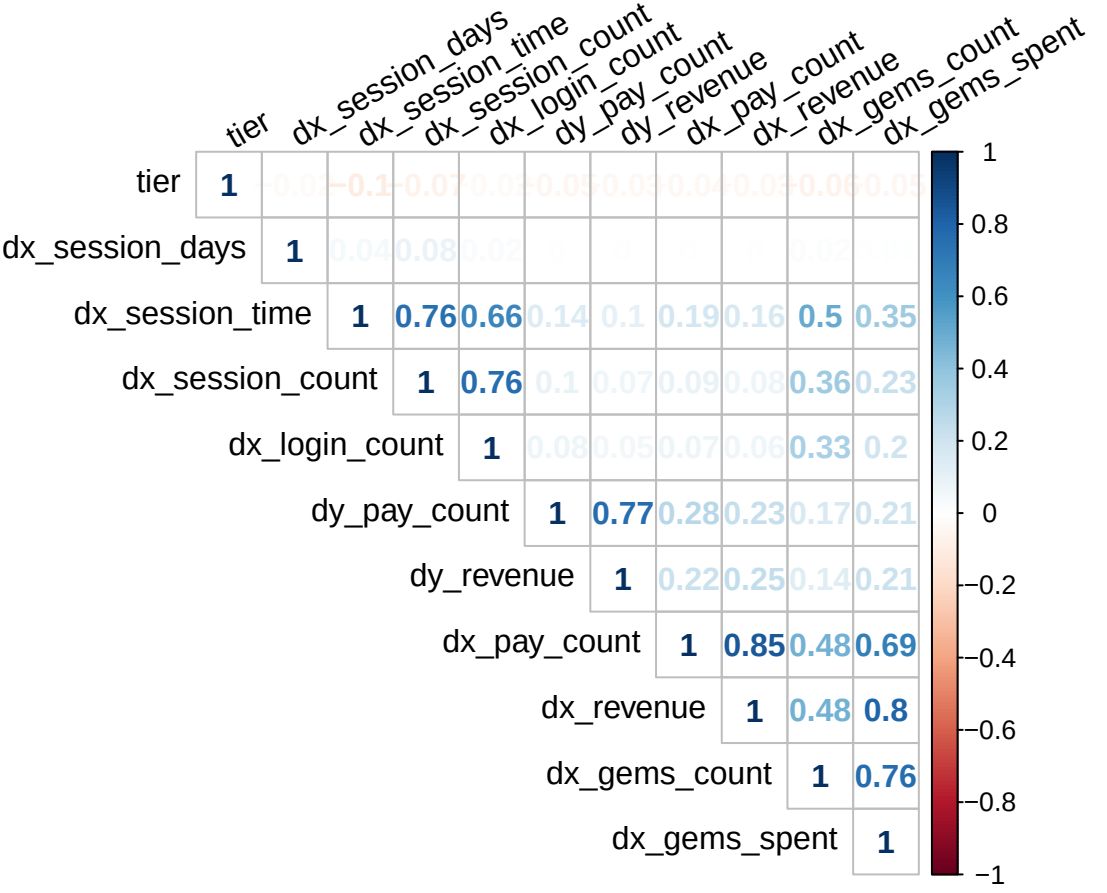
```
$dataFile
[1] "payer_model_DA_GP&iOS_mkt_2019-01-01_2019-03-31.rds"

$trainRegDate
[1] "2019-01-01" "2019-01-31"

$testRegDate
[1] "2019-03-01" "2019-03-31"
```

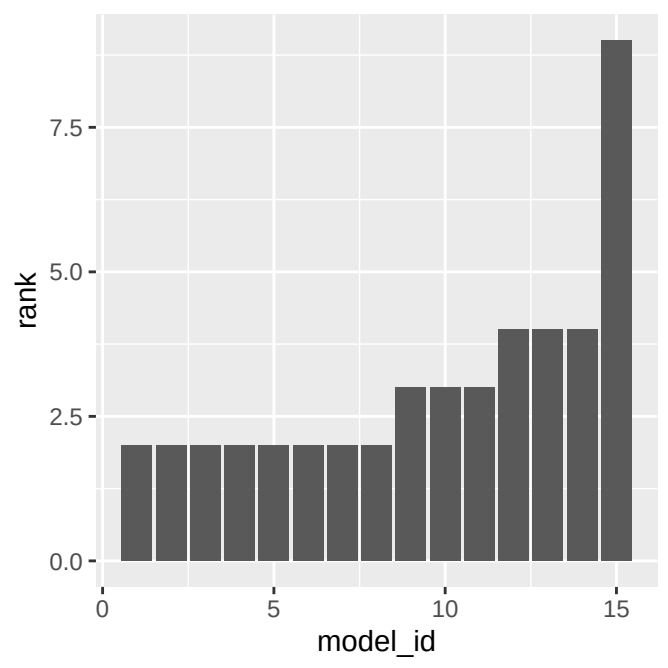
```
[1] "0_data/payer_model_DA_GP&iOS_mkt_2019-01-01_2019-03-31.rds"
```

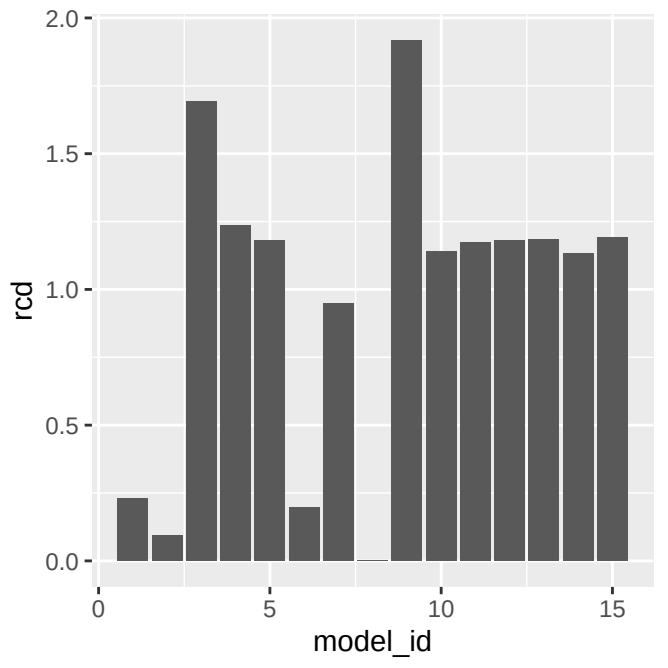

tier
dx_session_days
dx_session_time
dx_session_count
dx_login_count
dy_pay_count
dy_revenue
dx_pay_count
dx_revenue
dx_gems_count
dx_gems_spent
0.73 0.13 0.23 0.33 0.36 0.42 0.29 0.31 0.12 0.18
0.52 0.71 0.59 0.38 0.4 0.3 0.31 0.29 0.26
0.47 0.38 0.73 0.69 0.13 0.66
0.37 0.3 0.37 0.34 0.4 0.86
0.35 0.29 0.34 0.32 0.45 0.8
0.67 0.72 0.71 0.99
0.73 0.73 0.63 0.98
0.05
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

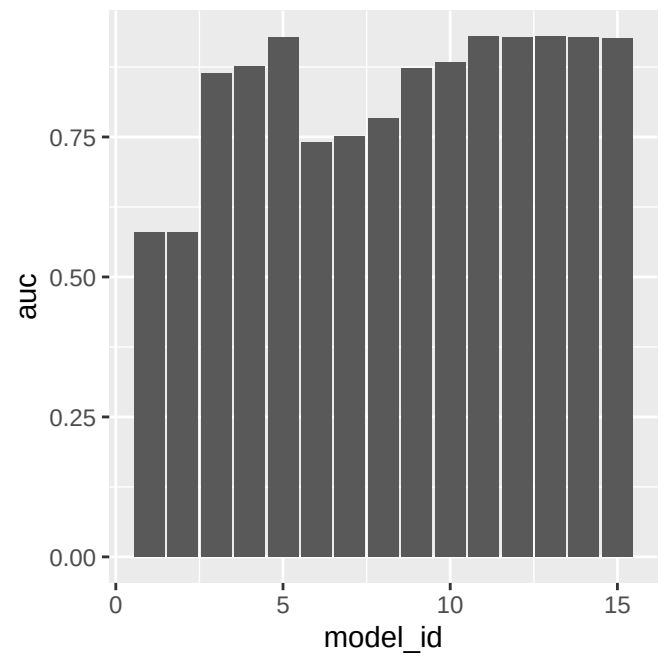

tier
dx_session_days
dx_session_time
dx_session_count
dx_login_count
dy_pay_count
dy_revenue
dx_pay_count
dx_revenue
dx_gems_count
dx_gems_spent
1
0.04 0.08 0.02 0 0 0 0 0.02 0.01
−0.1 −0.07 0.03 0.05 0.03 0.04 0.03 0.06 0.05
0.76 0.66 0.14 0.1 0.19 0.16 0.5 0.35
0.76 0.1 0.07 0.09 0.08 0.36 0.23
0.08 0.05 0.07 0.06 0.33 0.2
0.77 0.28 0.23 0.17 0.21
0.22 0.25 0.14 0.21
0.85 0.48 0.69
0.48 0.8
0.76
1
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

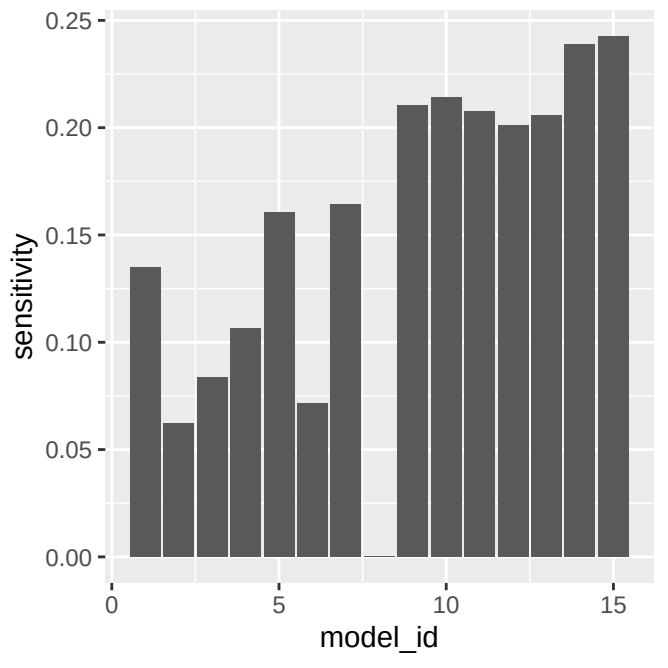
	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15
dx_revenue	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1
dx_pay_count	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
dx_login_count	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
dx_session_count	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1
dx_session_time	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1
dx_gems_count	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
dx_gems_spent	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1
tier	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1

	model_id	rank	auc	cut_off	rcd	sensitivity
1	1	2	0.5801201	0.078620482	0.23048669	0.13498623
2	2	2	0.5801178	0.944338017	0.09550046	0.06244261
3	3	2	0.8634584	0.042641134	1.69146006	0.08356290
4	4	2	0.8766328	0.049139189	1.23599633	0.10651974
5	5	2	0.9281521	0.085412310	1.17906336	0.16069789
6	6	2	0.7410610	0.145943877	0.19559229	0.07162534
7	7	2	0.7509478	0.007358032	0.94765840	0.16437098
8	8	2	0.7826341	0.999933893	0.00000000	0.00000000
9	9	3	0.8720839	0.033876536	1.91643710	0.21028466
10	10	3	0.8832331	0.055793261	1.13957759	0.21395776
11	11	3	0.9291643	0.076308575	1.17263545	0.20752984
12	12	4	0.9274954	0.078152990	1.18089991	0.20110193
13	13	4	0.9292691	0.076232792	1.18549128	0.20569330
14	14	4	0.9270993	0.092931179	1.13223140	0.23875115
15	15	9	0.9267396	0.096931795	1.19283747	0.24242424

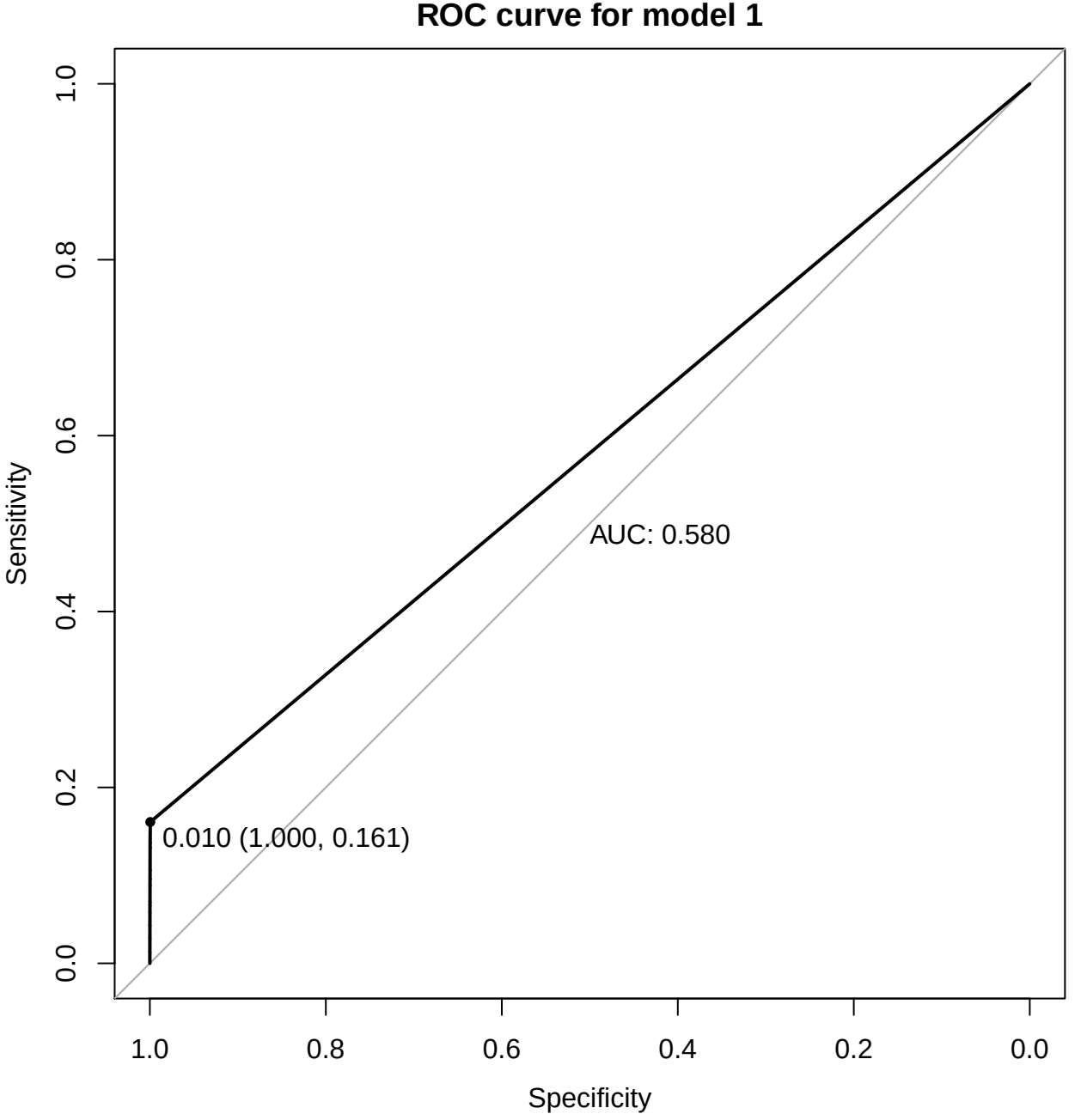

rank
7.5
5.0
2.5
0.0
0
5
10
15
model_id


rcd
2.0
1.5
1.0
0.5
0.0
0
5
10
15
model_id


auc
0.75
0.50
0.25
0.00
0
5
10
15
model_id


sensitivity
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
model_id

```
THIS SUMMARY IS OUTDATED
 WE FOUND OUT THAT WE NEED CROSS-VALIDATION TO GET ROBUST RESULTS
 Summary for DA:
 Based on correlation matrix, these variables should not be together
   dx_session_time, dx_session_count, dx_login_count
   dx_pay_count, dx_revenue
   dx_gems_spent, dx_gems_count
   dx_gems_spent, dx_revenue
   dx_gems_spent, dx_pay_count
 We will test 1-variable models
   dx_revenue ~ dx_pay_count (we choose revenue)
   dx_session_time > dx_session_count > dx_login_count (as expected)
   dx_gems_count > dx_gems_spent
 Check session features with dx_revenue
   dx_session_count ~> dx_session_time > dx_login_count (time is cont.)
 Check gem features with previous
   dx_gems_spent = dx_gems_count (spent is cont.)
   gems do not add anything and are correlated with revenue
 Check tier
   it bends the ROC curve in a weird way but improves the model
 Comparison with the full model shows that it has slightly better performance.
 It should be compared with cross-validation whether this difference is robust.
 Best model so far has: dx_revenue, dx_session_time, tier
```

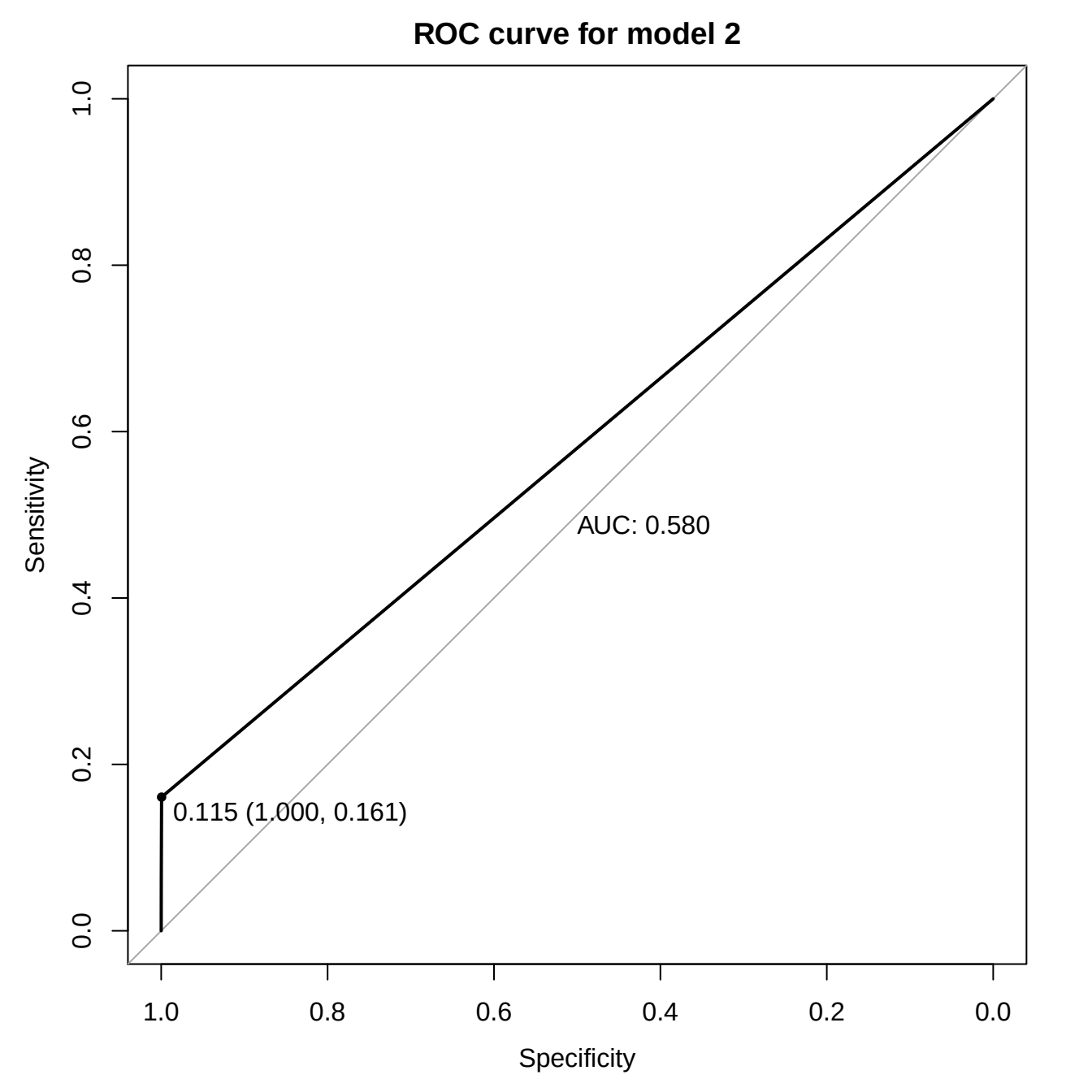

ROC curve for model 1
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
AUC: 0.580
0.010 (1.000, 0.161)
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−8.4904  −0.0739  −0.0739  −0.0739   3.4363

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) −5.90145    0.02705 −218.14   <2e−16 ***
dx_revenue   0.74185    0.02210   33.58   <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: 20791  on 501487  degrees of freedom
Residual deviance: 19128  on 501486  degrees of freedom
AIC: 19132

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.5801

$relativeCountDifference
[1] 0.2304867
```

```
$optimal_cut_off
[1] 0.07862048
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 338371    942
     TRUE     104    147

               Accuracy : 0.9969
                 95% CI : (0.9967, 0.9971)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 0.09794

                  Kappa : 0.2185

 Mcnemar's Test P-Value : < 2e-16

            Sensitivity : 0.1349862
            Specificity : 0.9996927
         Pos Pred Value : 0.5856574
         Neg Pred Value : 0.9972238
             Prevalence : 0.0032071
         Detection Rate : 0.0004329
   Detection Prevalence : 0.0007392
      Balanced Accuracy : 0.5673395

       'Positive' Class : TRUE
```

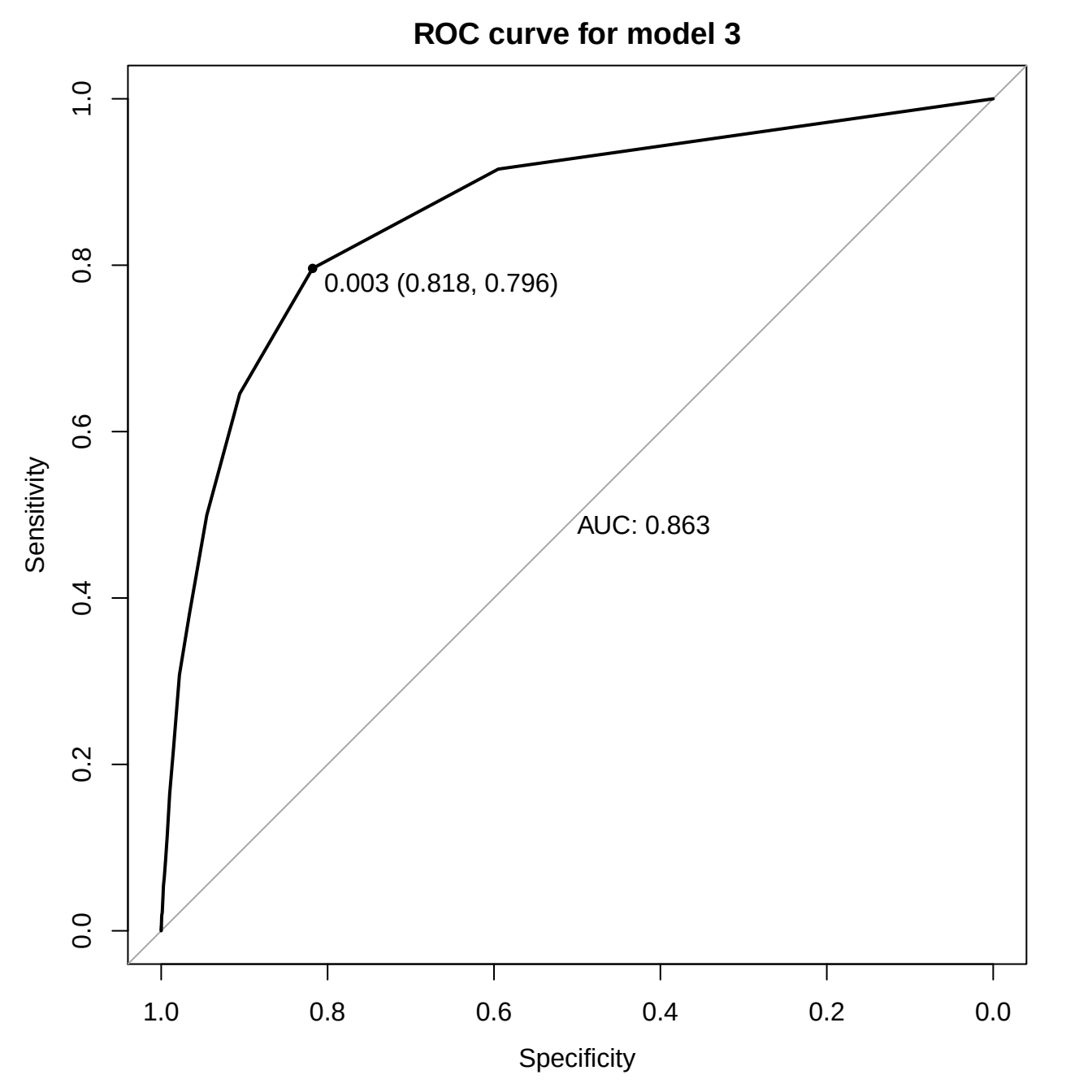

ROC curve for model 2
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
AUC: 0.580
0.115 (1.000, 0.161)
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−5.9324  −0.0729  −0.0729  −0.0729   3.4444

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  −5.92923    0.02744 −216.10   <2e−16 ***
dx_pay_count  4.70514    0.12775   36.83   <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: 20791  on 501487  degrees of freedom
Residual deviance: 18781  on 501486  degrees of freedom
AIC: 18785

Number of Fisher Scoring iterations: 8
```

```
$auc
Area under the curve: 0.5801

$relativeCountDifference
[1] 0.09550046
```

```
$optimal_cut_off
[1] 0.944338
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 338439   1021
     TRUE      36     68

               Accuracy : 0.9969
                 95% CI : (0.9967, 0.9971)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 0.1696

                  Kappa : 0.1135

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.0624426
            Specificity : 0.9998936
         Pos Pred Value : 0.6538462
         Neg Pred Value : 0.9969923
             Prevalence : 0.0032071
         Detection Rate : 0.0002003
   Detection Prevalence : 0.0003063
      Balanced Accuracy : 0.5311681

       'Positive' Class : TRUE
```

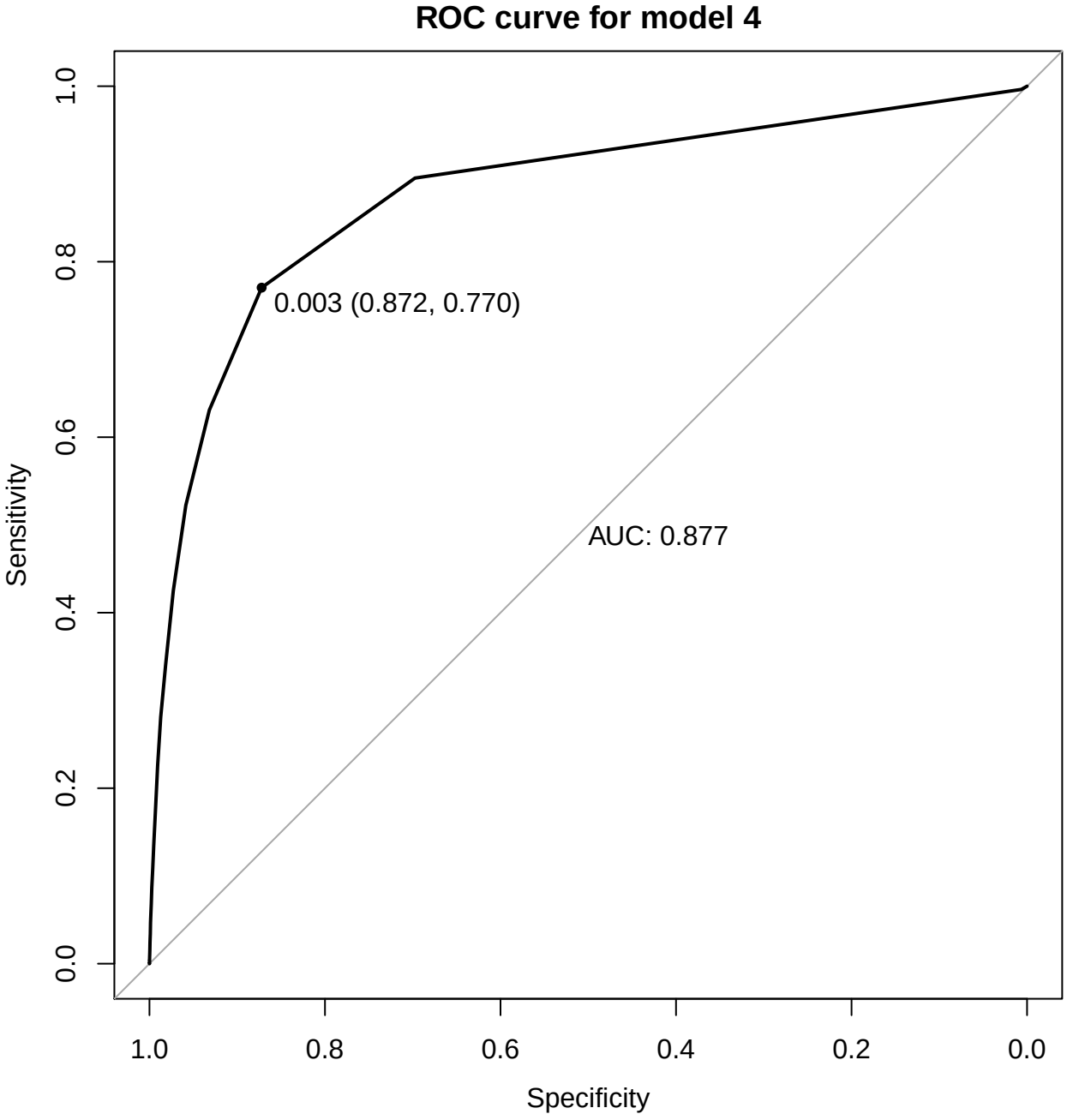

ROC curve for model 3
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.818, 0.796)
AUC: 0.863
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−6.8558  −0.0737  −0.0626  −0.0626   3.5314

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    −6.56018    0.03399 −193.00   <2e−16 ***
dx_login_count  0.32675    0.00597   54.73   <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: 20791  on 501487  degrees of freedom
Residual deviance: 18709  on 501486  degrees of freedom
AIC: 18713

Number of Fisher Scoring iterations: 8
```

```
$auc
Area under the curve: 0.8635

$relativeCountDifference
[1] 1.69146
```

```
$optimal_cut_off
[1] 0.04264113
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 336724    998
     TRUE    1751     91

               Accuracy : 0.9919
                 95% CI : (0.9916, 0.9922)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0583

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.083563
            Specificity : 0.994827
         Pos Pred Value : 0.049403
         Neg Pred Value : 0.997045
             Prevalence : 0.003207
         Detection Rate : 0.000268
   Detection Prevalence : 0.005425
      Balanced Accuracy : 0.539195

       'Positive' Class : TRUE
```

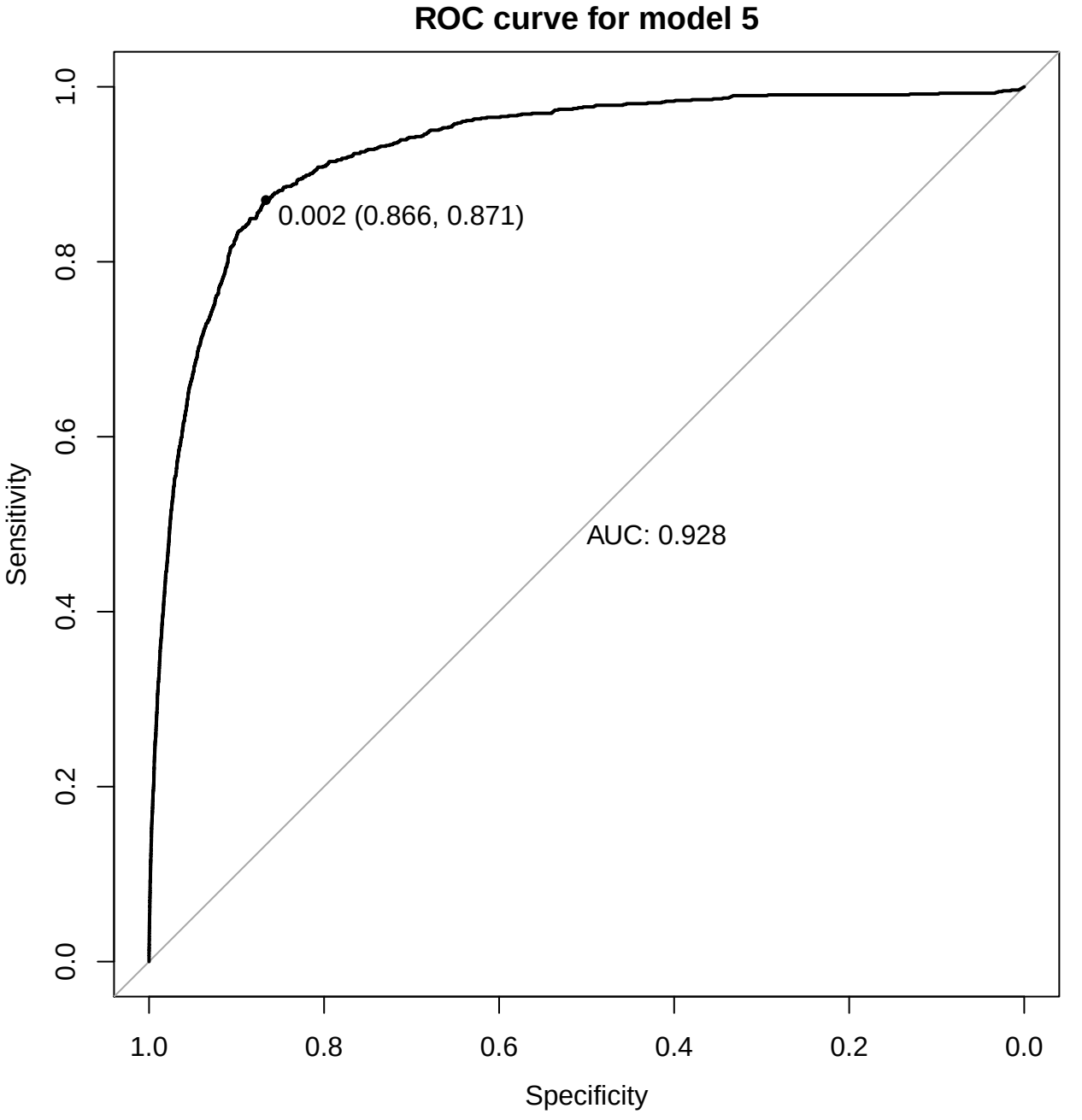

ROC curve for model 4
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.872, 0.770)
AUC: 0.877
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−3.5786  −0.0718  −0.0613  −0.0613   3.6316

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      −6.592734   0.033942 −194.24   <2e−16 ***
dx_session_count  0.316934   0.004909   64.56   <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: 20791  on 501487  degrees of freedom
Residual deviance: 17845  on 501486  degrees of freedom
AIC: 17849

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.8766

$relativeCountDifference
[1] 1.235996
```

```
$optimal_cut_off
[1] 0.04913919
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337245    973
     TRUE    1230    116

               Accuracy : 0.9935
                 95% CI : (0.9932, 0.9938)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0921

 Mcnemar's Test P-Value : 4.919e-08

            Sensitivity : 0.1065197
            Specificity : 0.9963661
         Pos Pred Value : 0.0861813
         Neg Pred Value : 0.9971232
             Prevalence : 0.0032071
         Detection Rate : 0.0003416
   Detection Prevalence : 0.0039639
      Balanced Accuracy : 0.5514429

       'Positive' Class : TRUE
```

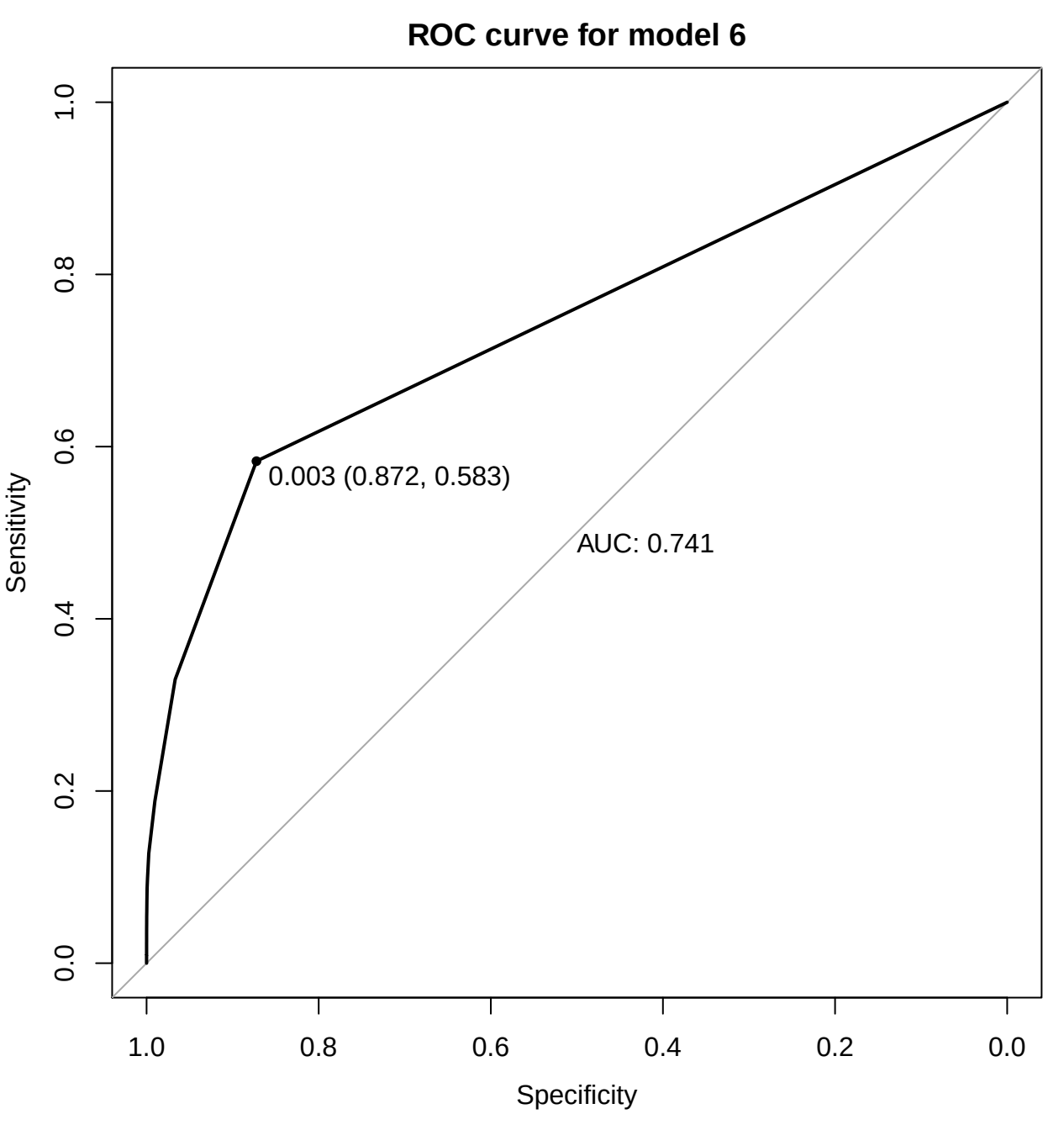

ROC curve for model 5
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.002 (0.866, 0.871)
AUC: 0.928
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−4.4975  −0.0599  −0.0550  −0.0531   3.6370

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     −6.6123601  0.0354110 −186.73   <2e−16 ***
dx_session_time  0.0165476  0.0002145   77.14   <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: 20791  on 501487  degrees of freedom
Residual deviance: 16131  on 501486  degrees of freedom
AIC: 16135

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.9282

$relativeCountDifference
[1] 1.179063
```

```
$optimal_cut_off
[1] 0.08541231
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337366    914
     TRUE    1109    175

               Accuracy : 0.994
                 95% CI : (0.9938, 0.9943)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1445

 Mcnemar's Test P-Value : 1.609e-05

            Sensitivity : 0.1606979
            Specificity : 0.9967235
         Pos Pred Value : 0.1362928
         Neg Pred Value : 0.9972981
             Prevalence : 0.0032071
         Detection Rate : 0.0005154
   Detection Prevalence : 0.0037813
      Balanced Accuracy : 0.5787107

       'Positive' Class : TRUE
```

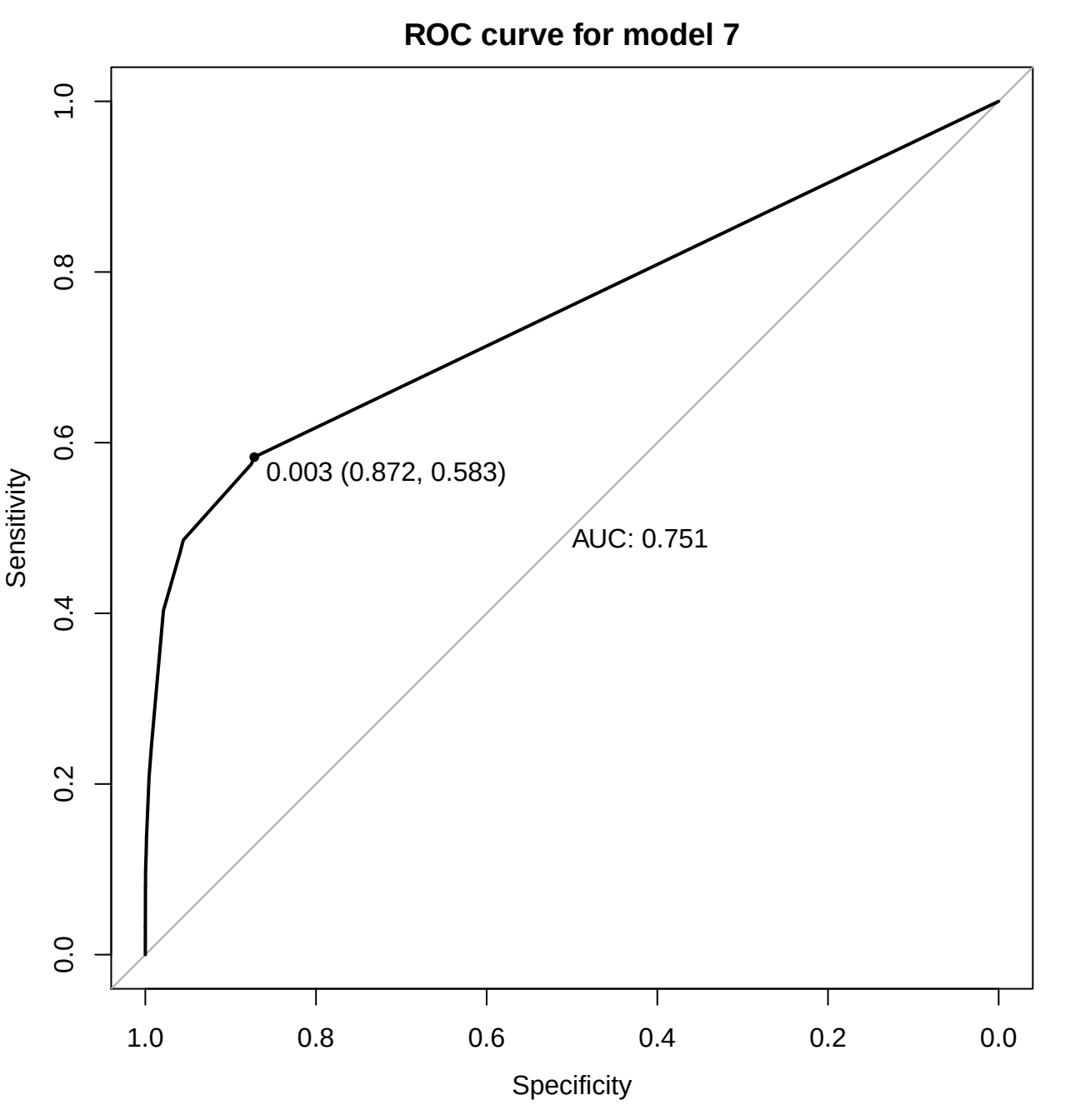

ROC curve for model 6
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.872, 0.583)
AUC: 0.741
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−7.3583  −0.0666  −0.0666  −0.0666   3.4962

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   −6.10944    0.02938 −207.95   <2e−16 ***
dx_gems_count  0.72135    0.01495   48.26   <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: 20791  on 501487  degrees of freedom
Residual deviance: 18306  on 501486  degrees of freedom
AIC: 18310

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.7411

$relativeCountDifference
[1] 0.1955923
```

```
$optimal_cut_off
[1] 0.1459439
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 338340   1011
     TRUE     135     78

               Accuracy : 0.9966
                 95% CI : (0.9964, 0.9968)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 0.9586

                  Kappa : 0.1189

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.0716253
            Specificity : 0.9996012
         Pos Pred Value : 0.3661972
         Neg Pred Value : 0.9970208
             Prevalence : 0.0032071
         Detection Rate : 0.0002297
   Detection Prevalence : 0.0006273
      Balanced Accuracy : 0.5356132

       'Positive' Class : TRUE
```

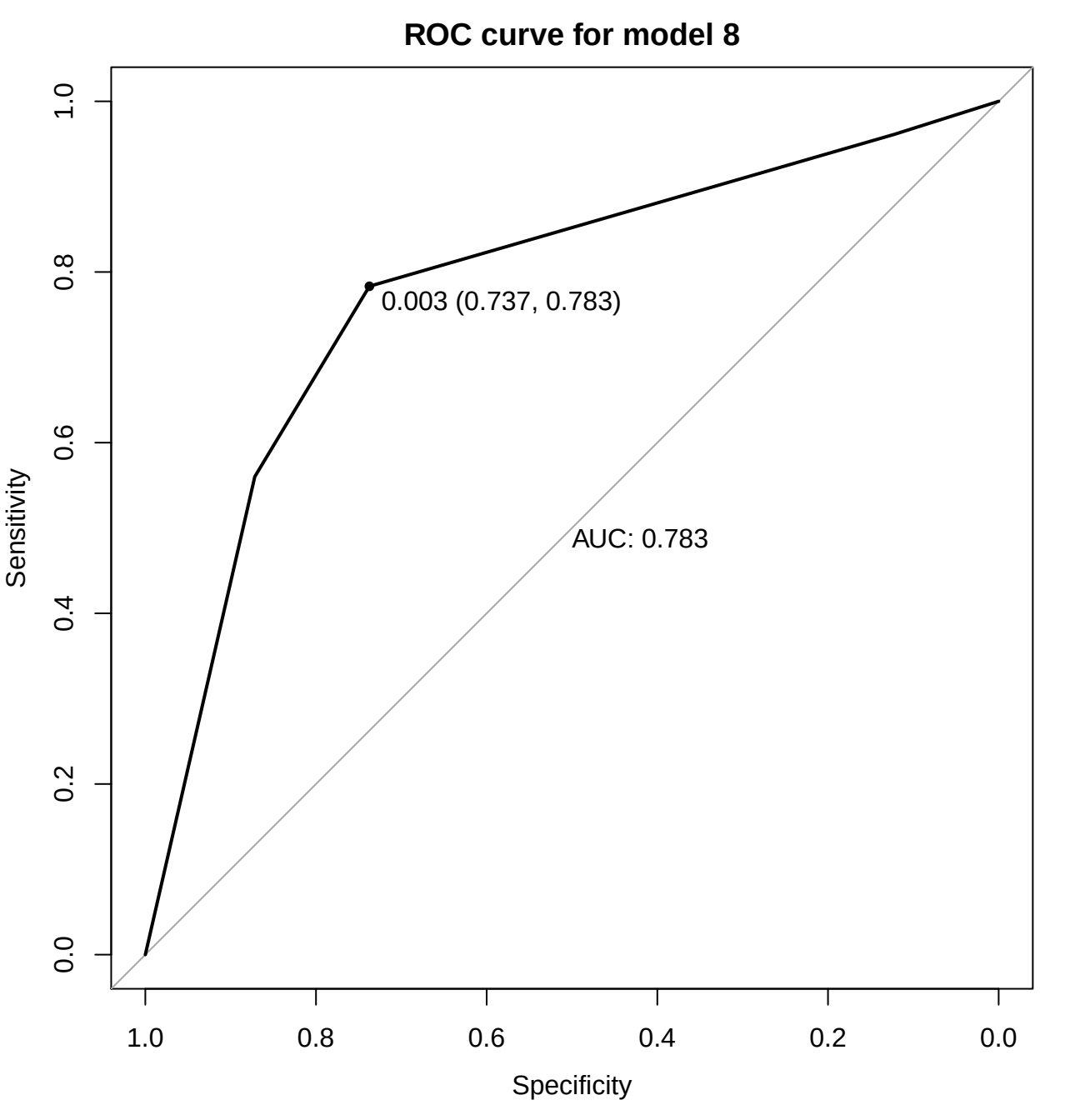

ROC curve for model 7
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.872, 0.583)
AUC: 0.751
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−8.4904  −0.0713  −0.0713  −0.0713   3.4571

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   −5.9733547  0.0277099  −215.6   <2e−16 ***
dx_gems_spent  0.0288496  0.0008036    35.9   <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: 20791  on 501487  degrees of freedom
Residual deviance: 18761  on 501486  degrees of freedom
AIC: 18765

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.7509

$relativeCountDifference
[1] 0.9476584
```

```
$optimal_cut_off
[1] 0.007358032
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337622    910
     TRUE     853    179

               Accuracy : 0.9948
                 95% CI : (0.9946, 0.995)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1.0000

                  Kappa : 0.1662

 Mcnemar's Test P-Value : 0.1823

            Sensitivity : 0.1643710
            Specificity : 0.9974799
         Pos Pred Value : 0.1734496
         Neg Pred Value : 0.9973119
             Prevalence : 0.0032071
         Detection Rate : 0.0005271
   Detection Prevalence : 0.0030392
      Balanced Accuracy : 0.5809254

       'Positive' Class : TRUE
```

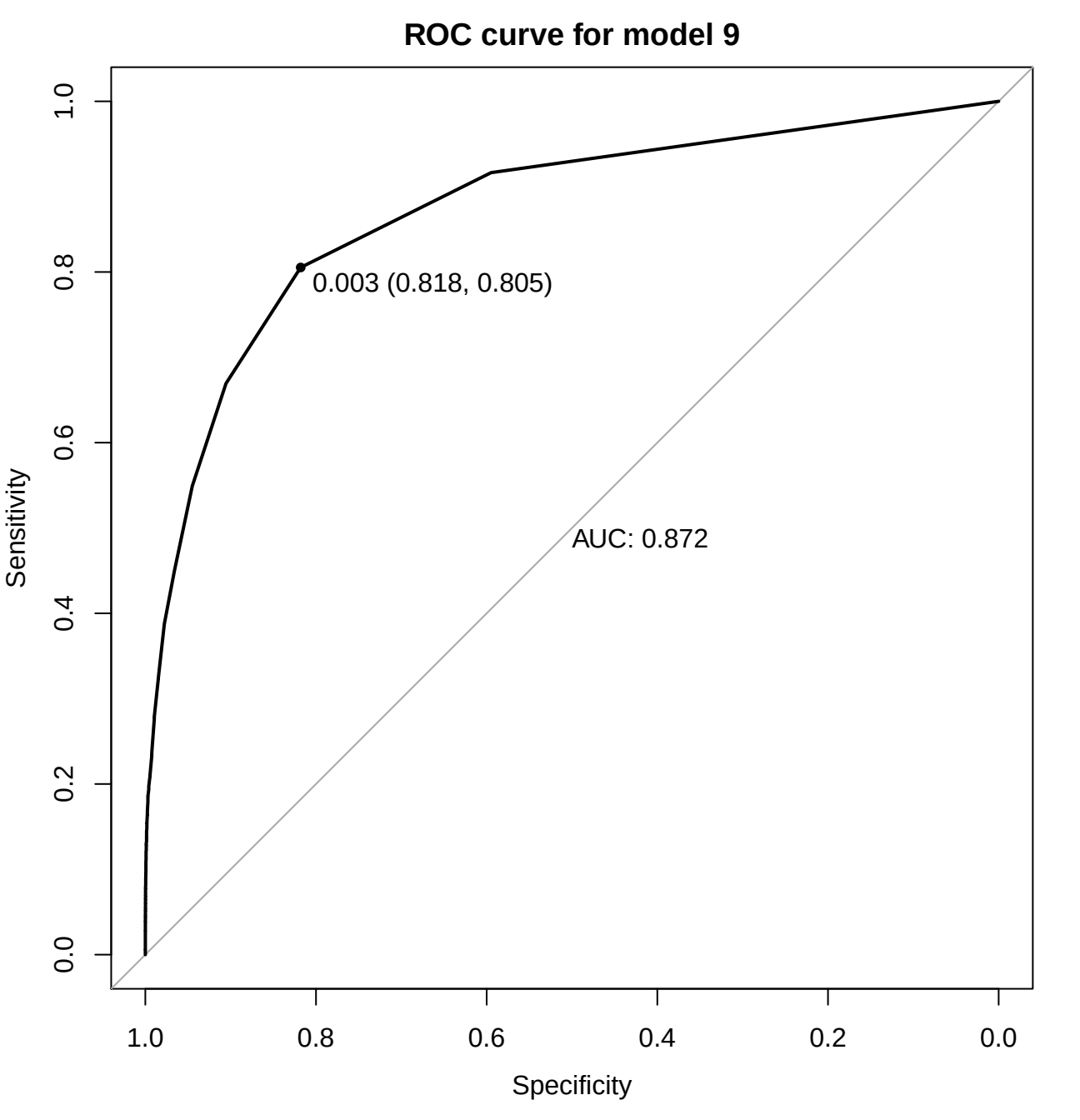

ROC curve for model 8
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.737, 0.783)
AUC: 0.783
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−0.1736  −0.0478  −0.0478  −0.0478   4.0173

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) −2.89339    0.05351  −54.07   <2e−16 ***
tier        −1.29393    0.02870  −45.09   <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: 20791  on 501487  degrees of freedom
Residual deviance: 18397  on 501486  degrees of freedom
AIC: 18401

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.7826

$relativeCountDifference
[1] 0
```

```
$optimal_cut_off
[1] 0.9999339
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 338475   1089
     TRUE       0      0

               Accuracy : 0.9968
                 95% CI : (0.9966, 0.997)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 0.5081

                  Kappa : 0

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.000000
            Specificity : 1.000000
         Pos Pred Value :      NaN
         Neg Pred Value : 0.996793
             Prevalence : 0.003207
         Detection Rate : 0.000000
   Detection Prevalence : 0.000000
      Balanced Accuracy : 0.500000

       'Positive' Class : TRUE
```

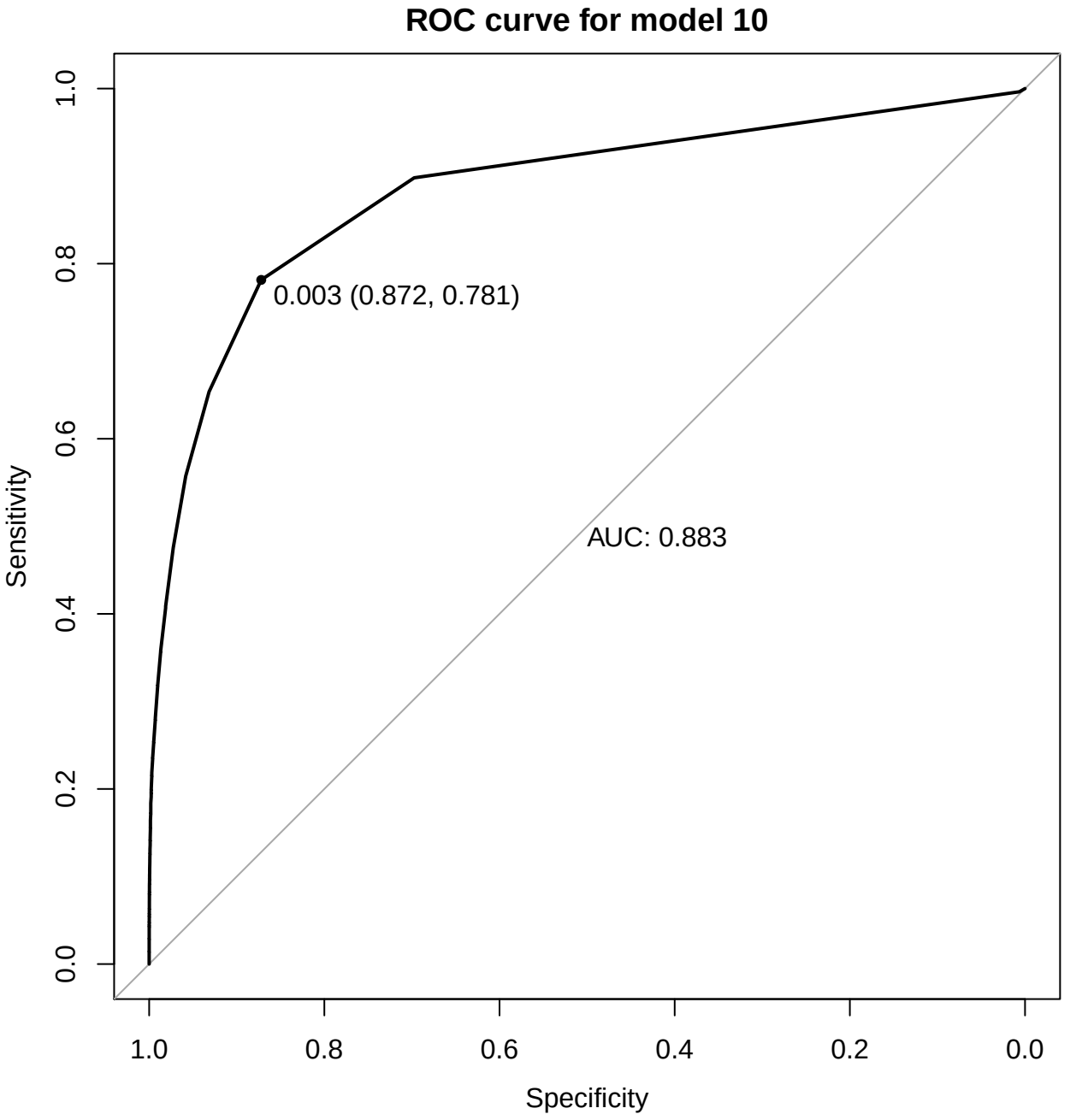

ROC curve for model 9
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.818, 0.805)
AUC: 0.872
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−8.4904  −0.0704  −0.0603  −0.0603   3.5531

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    −6.620348   0.035304 −187.52   <2e−16 ***
dx_revenue      0.597167   0.020054   29.78   <2e−16 ***
dx_login_count  0.310015   0.006241   49.67   <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: 20791  on 501487  degrees of freedom
Residual deviance: 17468  on 501485  degrees of freedom
AIC: 17474

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.8721

$relativeCountDifference
[1] 1.916437
```

```
$optimal_cut_off
[1] 0.03387654
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 336617    860
     TRUE    1858    229

               Accuracy : 0.992
                 95% CI : (0.9917, 0.9923)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1406

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.2102847
            Specificity : 0.9945107
         Pos Pred Value : 0.1097269
         Neg Pred Value : 0.9974517
             Prevalence : 0.0032071
         Detection Rate : 0.0006744
   Detection Prevalence : 0.0061461
      Balanced Accuracy : 0.6023977

       'Positive' Class : TRUE
```

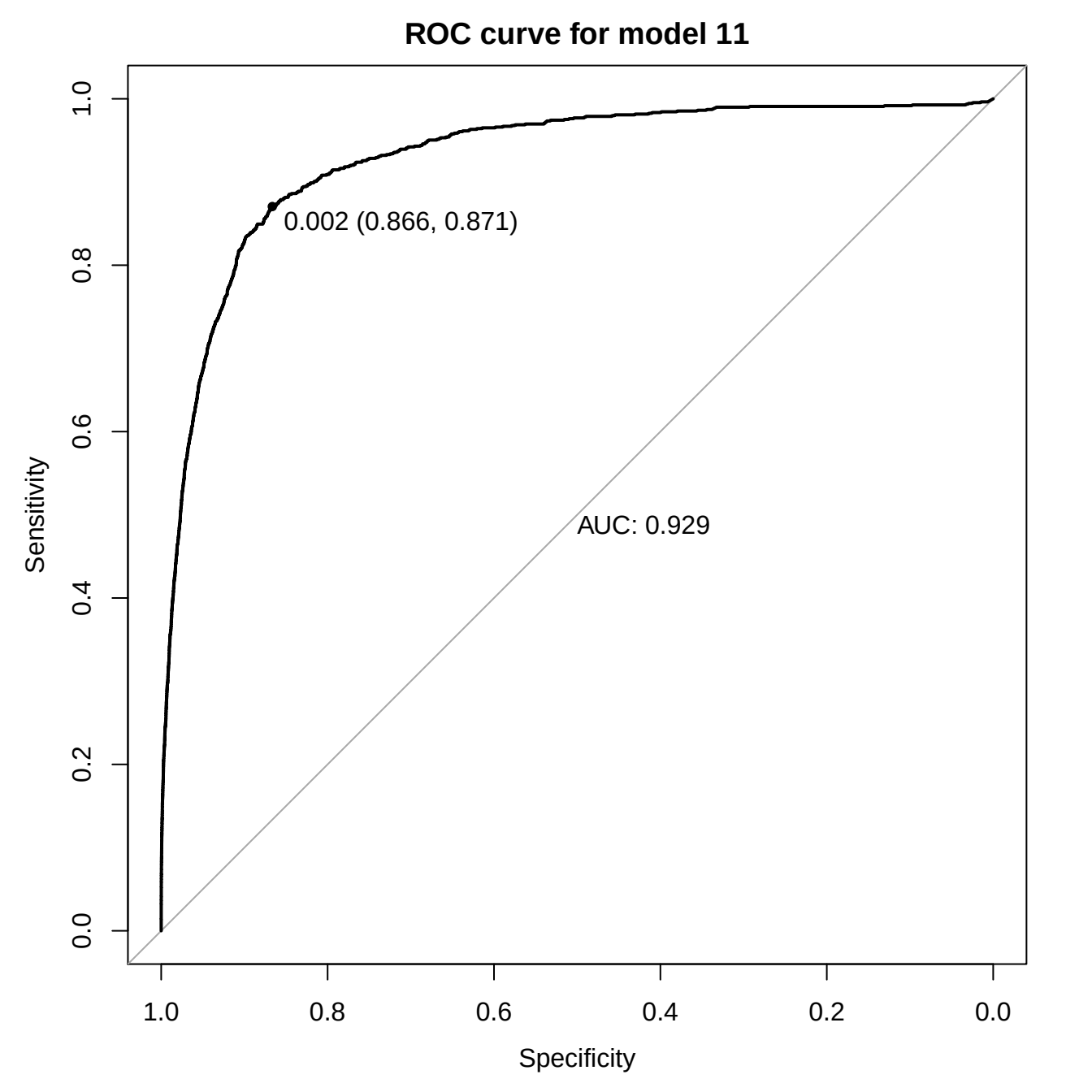

ROC curve for model 10
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.872, 0.781)
AUC: 0.883
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−8.4904  −0.0689  −0.0594  −0.0594   3.6445

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      −6.639861   0.035040 −189.49   <2e−16 ***
dx_revenue        0.535200   0.019293   27.74   <2e−16 ***
dx_session_count  0.299287   0.005106   58.61   <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: 20791  on 501487  degrees of freedom
Residual deviance: 16795  on 501485  degrees of freedom
AIC: 16801

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.8832

$relativeCountDifference
[1] 1.139578
```

```
$optimal_cut_off
[1] 0.05579326
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337467    856
     TRUE    1008    233

               Accuracy : 0.9945
                 95% CI : (0.9943, 0.9948)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1.0000000

                  Kappa : 0.1973

 Mcnemar's Test P-Value : 0.0004697

            Sensitivity : 0.2139578
            Specificity : 0.9970219
         Pos Pred Value : 0.1877518
         Neg Pred Value : 0.9974699
             Prevalence : 0.0032071
         Detection Rate : 0.0006862
   Detection Prevalence : 0.0036547
      Balanced Accuracy : 0.6054898

       'Positive' Class : TRUE
```

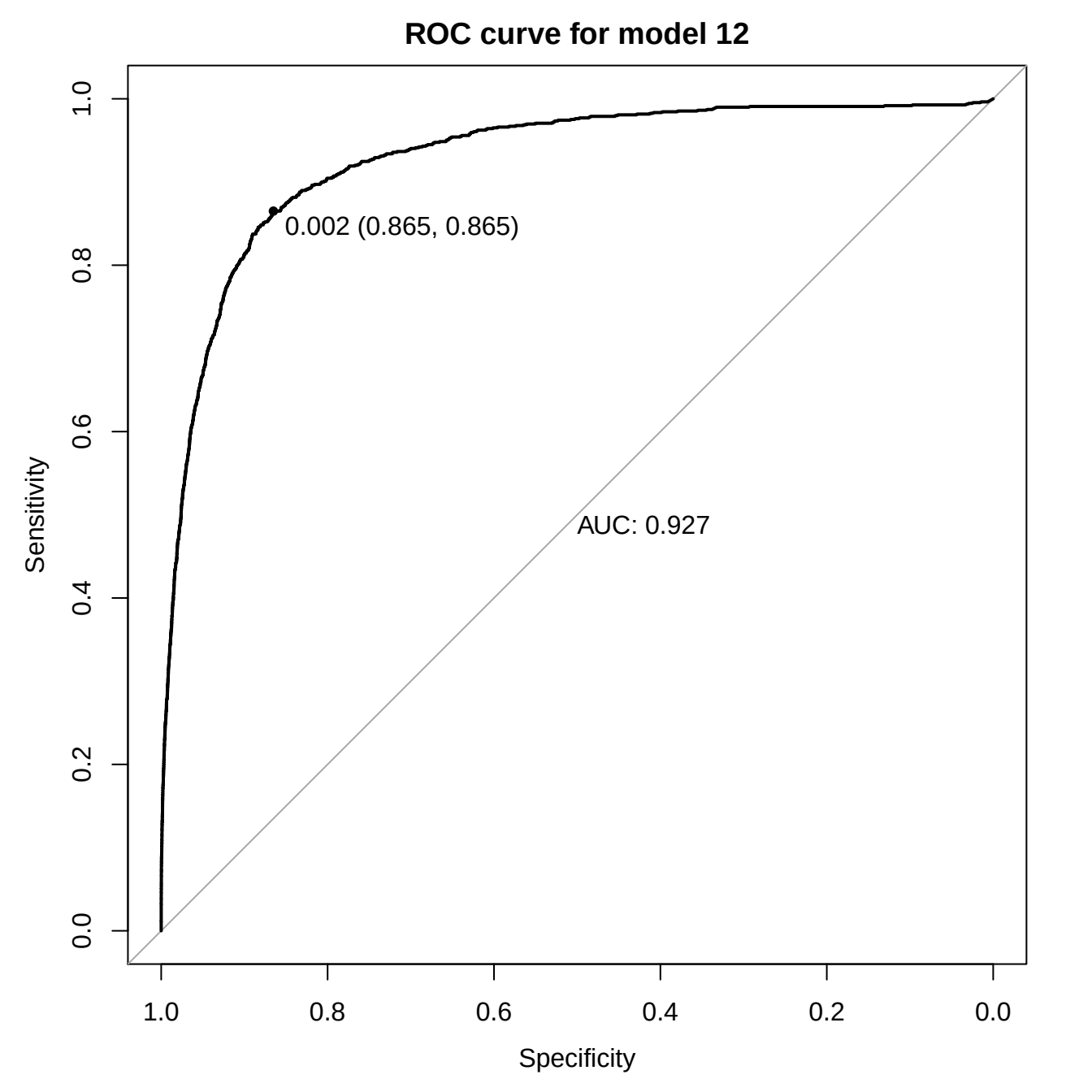

ROC curve for model 11
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.002 (0.866, 0.871)
AUC: 0.929
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−6.4365  −0.0596  −0.0551  −0.0533   3.6343

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     −6.6025613  0.0355739 −185.60   <2e−16 ***
dx_revenue       0.3278033  0.0187887   17.45   <2e−16 ***
dx_session_time  0.0155640  0.0002275   68.41   <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: 20791  on 501487  degrees of freedom
Residual deviance: 15762  on 501485  degrees of freedom
AIC: 15768

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.9292

$relativeCountDifference
[1] 1.172635
```

```
$optimal_cut_off
[1] 0.07630857
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337424    863
     TRUE    1051    226

               Accuracy : 0.9944
                 95% CI : (0.9941, 0.9946)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1882

 Mcnemar's Test P-Value : 1.917e-05

            Sensitivity : 0.2075298
            Specificity : 0.9968949
         Pos Pred Value : 0.1769773
         Neg Pred Value : 0.9974489
             Prevalence : 0.0032071
         Detection Rate : 0.0006656
   Detection Prevalence : 0.0037607
      Balanced Accuracy : 0.6022124

       'Positive' Class : TRUE
```

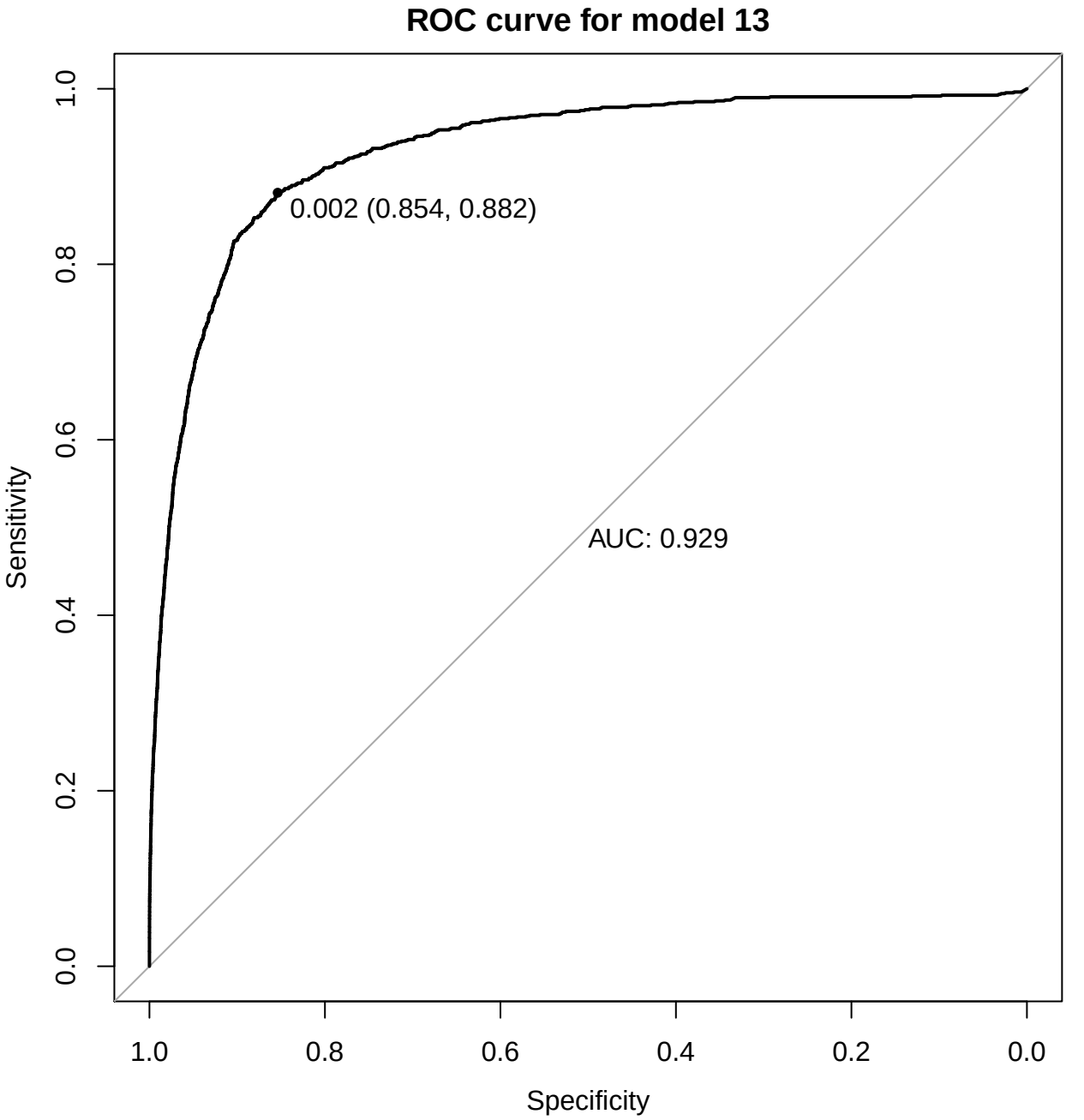

ROC curve for model 12
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
0.002 (0.865, 0.865)
AUC: 0.927

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−6.1345  −0.0583  −0.0541  −0.0524   3.6421

Coefficients:
                  Estimate Std. Error  z value Pr(>|z|)
(Intercept)     −6.6312240  0.0360418 −183.987   <2e−16 ***
dx_revenue       0.2058785  0.0220331    9.344   <2e−16 ***
dx_session_time  0.0145459  0.0002512   57.898   <2e−16 ***
dx_gems_count    0.1882357  0.0187400   10.045   <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: 20791  on 501487  degrees of freedom
Residual deviance: 15663  on 501484  degrees of freedom
AIC: 15671

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.9275

$relativeCountDifference
[1] 1.1809
```

```
$optimal_cut_off
[1] 0.07815299
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337408    870
     TRUE    1067    219

               Accuracy : 0.9943
                 95% CI : (0.994, 0.9945)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1816

 Mcnemar's Test P-Value : 8.452e-06

            Sensitivity : 0.2011019
            Specificity : 0.9968476
         Pos Pred Value : 0.1702955
         Neg Pred Value : 0.9974282
             Prevalence : 0.0032071
         Detection Rate : 0.0006449
   Detection Prevalence : 0.0037872
      Balanced Accuracy : 0.5989748

       'Positive' Class : TRUE
```

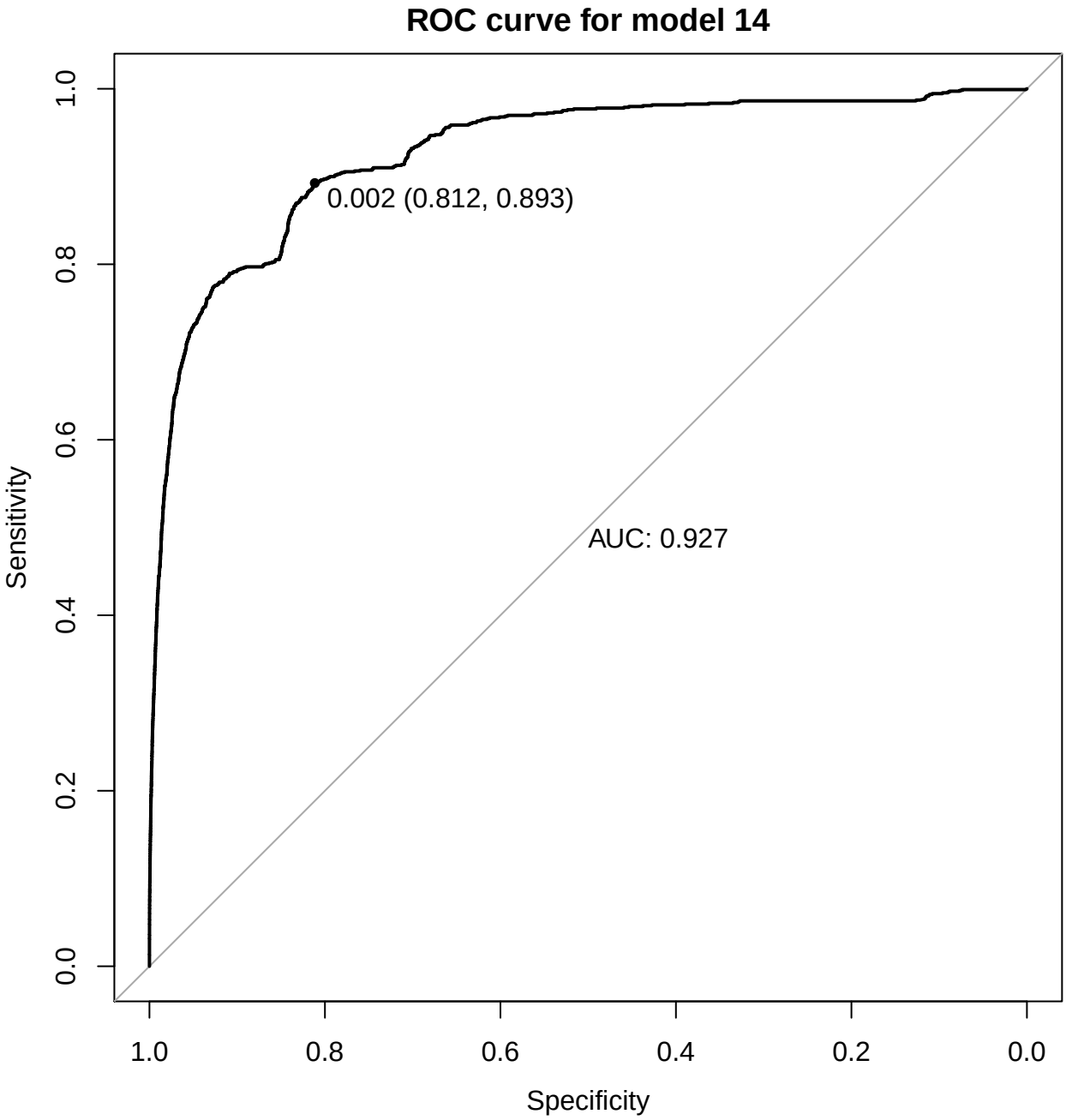

ROC curve for model 13
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.002 (0.854, 0.882)
AUC: 0.929
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−6.7758  −0.0594  −0.0549  −0.0531   3.6355

Coefficients:
                  Estimate Std. Error  z value Pr(>|z|)
(Intercept)     −6.6072095  0.0356510 −185.330  < 2e−16 ***
dx_revenue       0.2408844  0.0294707    8.174 2.99e−16 ***
dx_session_time  0.0152858  0.0002386   64.074  < 2e−16 ***
dx_gems_spent    0.0040191  0.0010412    3.860 0.000113 ***
−−−
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 20791  on 501487  degrees of freedom
Residual deviance: 15746  on 501484  degrees of freedom
AIC: 15754

Number of Fisher Scoring iterations: 9
```

```
$auc
Area under the curve: 0.9293

$relativeCountDifference
[1] 1.185491
```

```
$optimal_cut_off
[1] 0.07623279
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337408    865
     TRUE    1067    224

               Accuracy : 0.9943
                 95% CI : (0.9941, 0.9946)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1854

 Mcnemar's Test P-Value : 4.81e-06

            Sensitivity : 0.2056933
            Specificity : 0.9968476
         Pos Pred Value : 0.1735089
         Neg Pred Value : 0.9974429
             Prevalence : 0.0032071
         Detection Rate : 0.0006597
   Detection Prevalence : 0.0038019
      Balanced Accuracy : 0.6012705

       'Positive' Class : TRUE
```

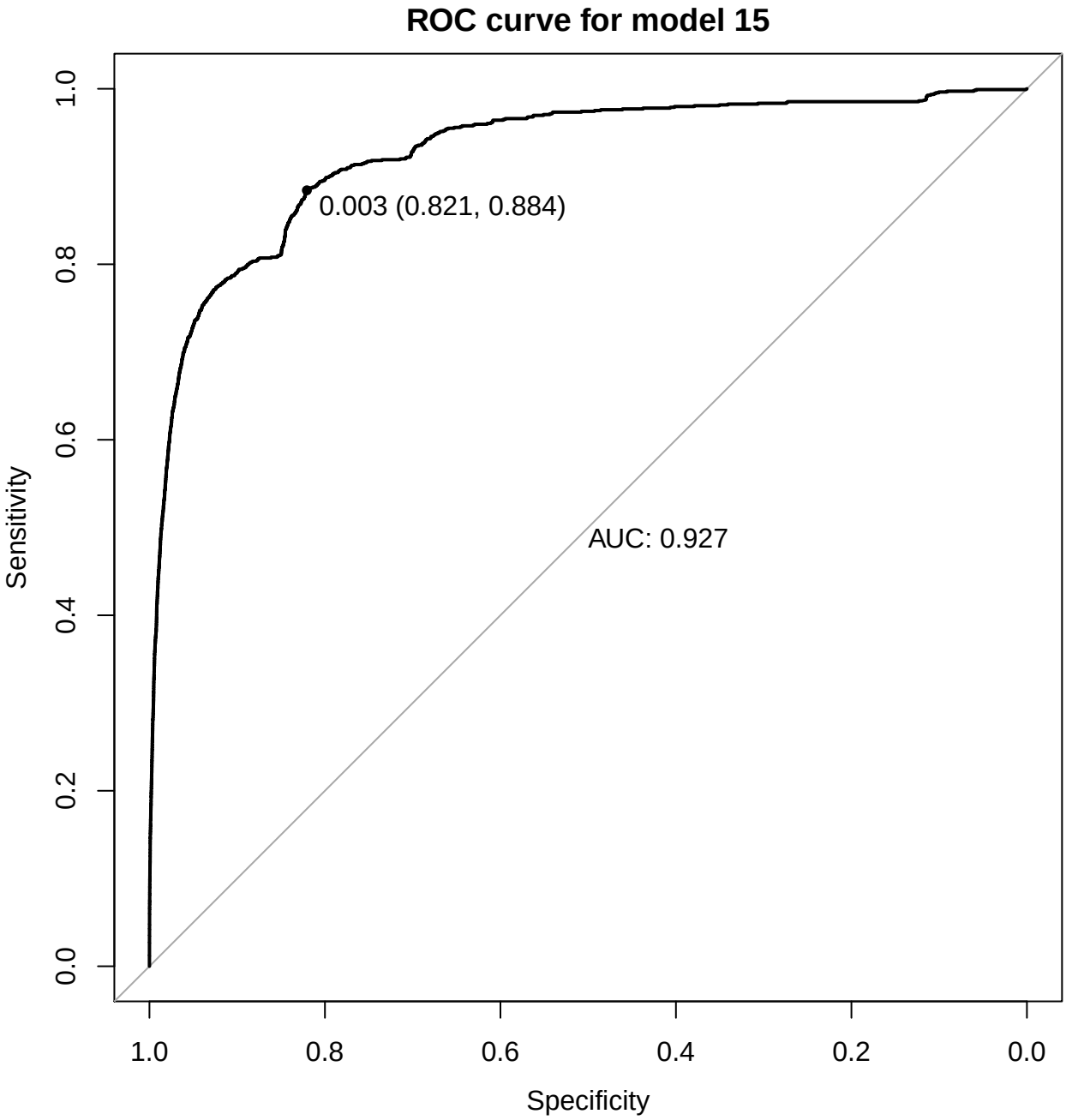

ROC curve for model 14
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.002 (0.812, 0.893)
AUC: 0.927
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−5.5755  −0.0488  −0.0358  −0.0343   4.0919

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     −4.0020543  0.0639137  −62.62   <2e−16 ***
dx_revenue       0.2333907  0.0180199   12.95   <2e−16 ***
dx_session_time  0.0148375  0.0002391   62.05   <2e−16 ***
tier            −1.1552388  0.0313291  −36.87   <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: 20791  on 501487  degrees of freedom
Residual deviance: 14197  on 501484  degrees of freedom
AIC: 14205

Number of Fisher Scoring iterations: 10
```

```
$auc
Area under the curve: 0.9271

$relativeCountDifference
[1] 1.132231
```

```
$optimal_cut_off
[1] 0.09293118
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337502    829
     TRUE     973    260

               Accuracy : 0.9947
                 95% CI : (0.9944, 0.9949)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1.0000000

                  Kappa : 0.2213

 Mcnemar's Test P-Value : 0.0007553

            Sensitivity : 0.2387511
            Specificity : 0.9971253
         Pos Pred Value : 0.2108678
         Neg Pred Value : 0.9975497
             Prevalence : 0.0032071
         Detection Rate : 0.0007657
   Detection Prevalence : 0.0036311
      Balanced Accuracy : 0.6179382

       'Positive' Class : TRUE
```


ROC curve for model 15
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.003 (0.821, 0.884)
AUC: 0.927
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

```
Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
−5.1345  −0.0499  −0.0347  −0.0330   4.1278

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)      −4.1598825  0.0655416 −63.469  < 2e−16 ***
dx_revenue       −0.0577210  0.0306323  −1.884 0.059522 .
dx_pay_count      2.0206826  0.1885860  10.715  < 2e−16 ***
dx_login_count    0.0631714  0.0106041   5.957 2.57e−09 ***
dx_session_count  0.0567058  0.0103062   5.502 3.75e−08 ***
dx_session_time   0.0106745  0.0004178  25.550  < 2e−16 ***
dx_gems_count     0.1929735  0.0273877   7.046 1.84e−12 ***
dx_gems_spent    −0.0050308  0.0013322  −3.776 0.000159 ***
tier             −1.1650272  0.0317373 −36.708  < 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: 20791  on 501487  degrees of freedom
Residual deviance: 13887  on 501479  degrees of freedom
AIC: 13905

Number of Fisher Scoring iterations: 10
```

```
$auc
Area under the curve: 0.9267

$relativeCountDifference
[1] 1.192837
```

```
$optimal_cut_off
[1] 0.09693179
```

```
$confMatrix
Confusion Matrix and Statistics

          Reference
Prediction  FALSE   TRUE
     FALSE 337440    825
     TRUE    1035    264

               Accuracy : 0.9945
                 95% CI : (0.9943, 0.9948)
    No Information Rate : 0.9968
    P-Value [Acc > NIR] : 1

                  Kappa : 0.2184

 Mcnemar's Test P-Value : 1.259e-06

            Sensitivity : 0.2424242
            Specificity : 0.9969422
         Pos Pred Value : 0.2032333
         Neg Pred Value : 0.9975611
             Prevalence : 0.0032071
         Detection Rate : 0.0007775
   Detection Prevalence : 0.0038255
      Balanced Accuracy : 0.6196832

       'Positive' Class : TRUE
```