Estudo de Caso - Diagnóstico de Câncer de Mama

Bruno Crotman

25/06/2019

1 Origem dos dados

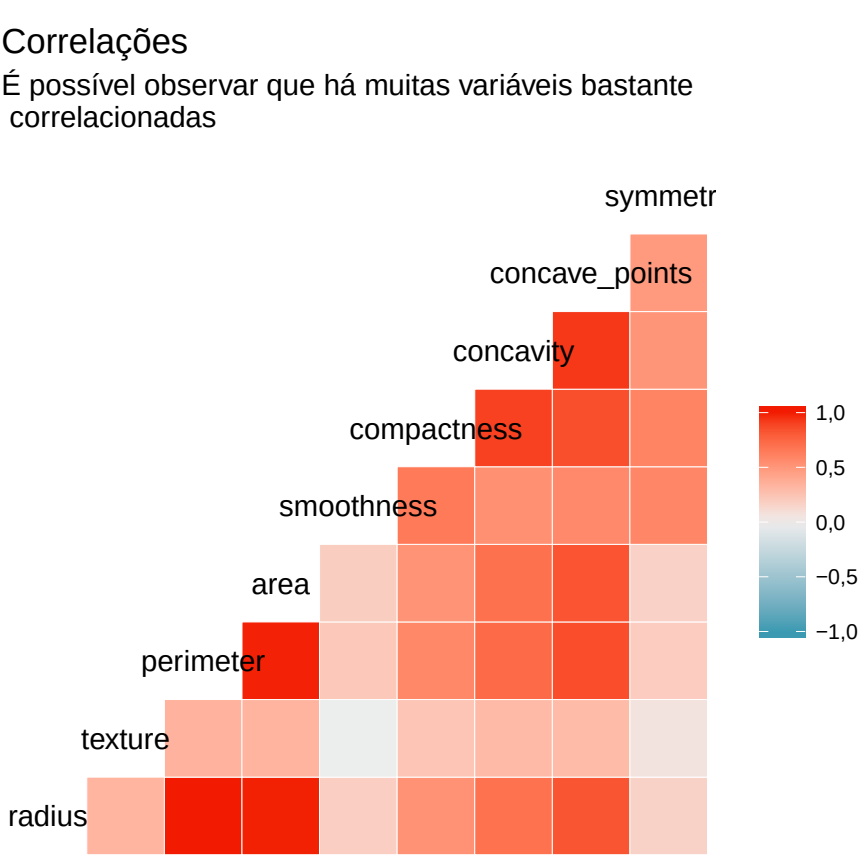
Os dados utilizados vieram do site UCI Machine Learning Repository - Breast Cancer Wisconsin (Diagnostic) Data Set

A base contém 569, com 31 variáveis, sendo uma delas o diagnóstico de benigno ou maligno.

2 Conjuntos Treino e Teste

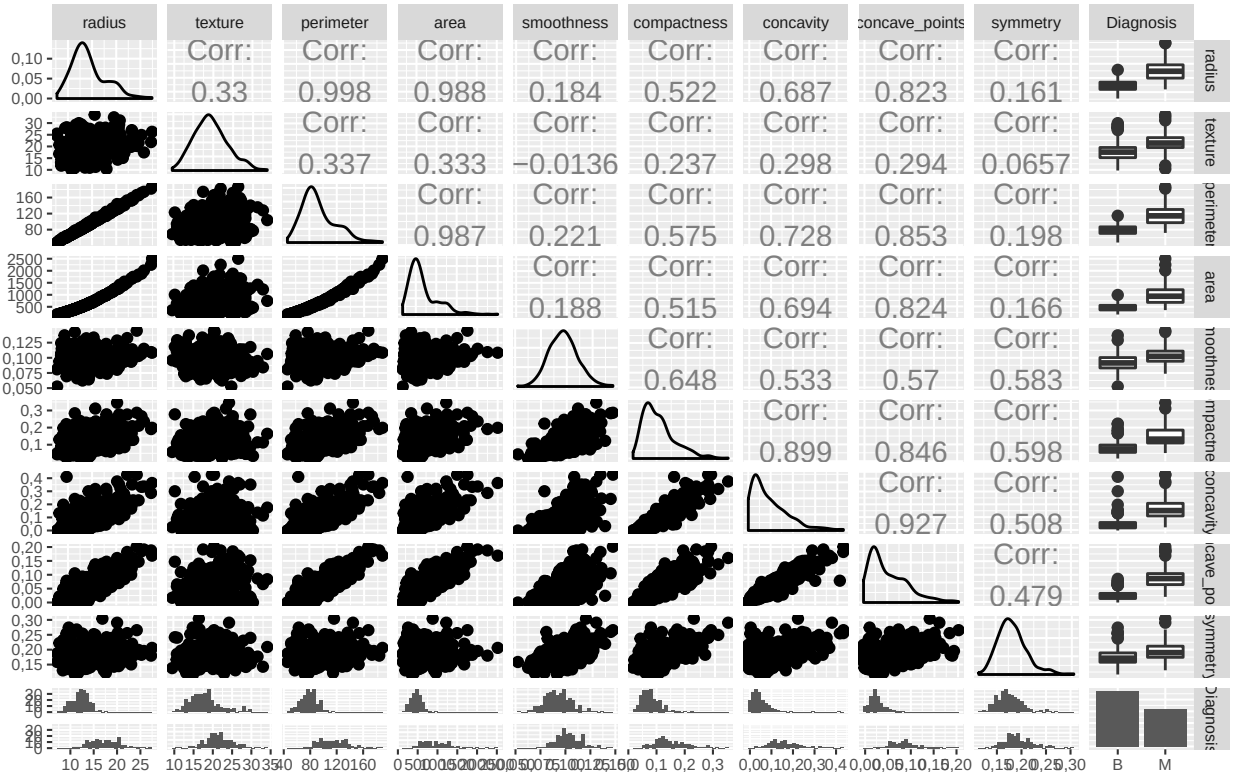
Já antes o início da análise, a base foi dividida aleatoriamente em dados de teste e dados de treino. Os dados de teste não foram tocados durante nenhuma parte da análise. Apenas na avaliação final dos modelos.

Os dados de treino e teste constituem, respectivamente, 70% e 30% do total.



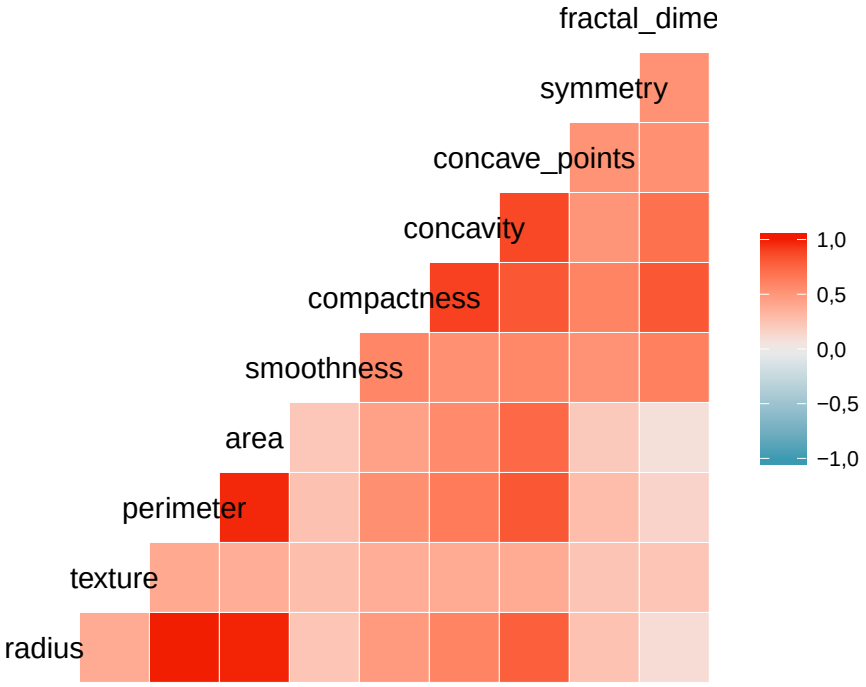
Resumo das distribuições

É possível observar que o valor da maioria das variáveis se mostra maior para os resultados malignos



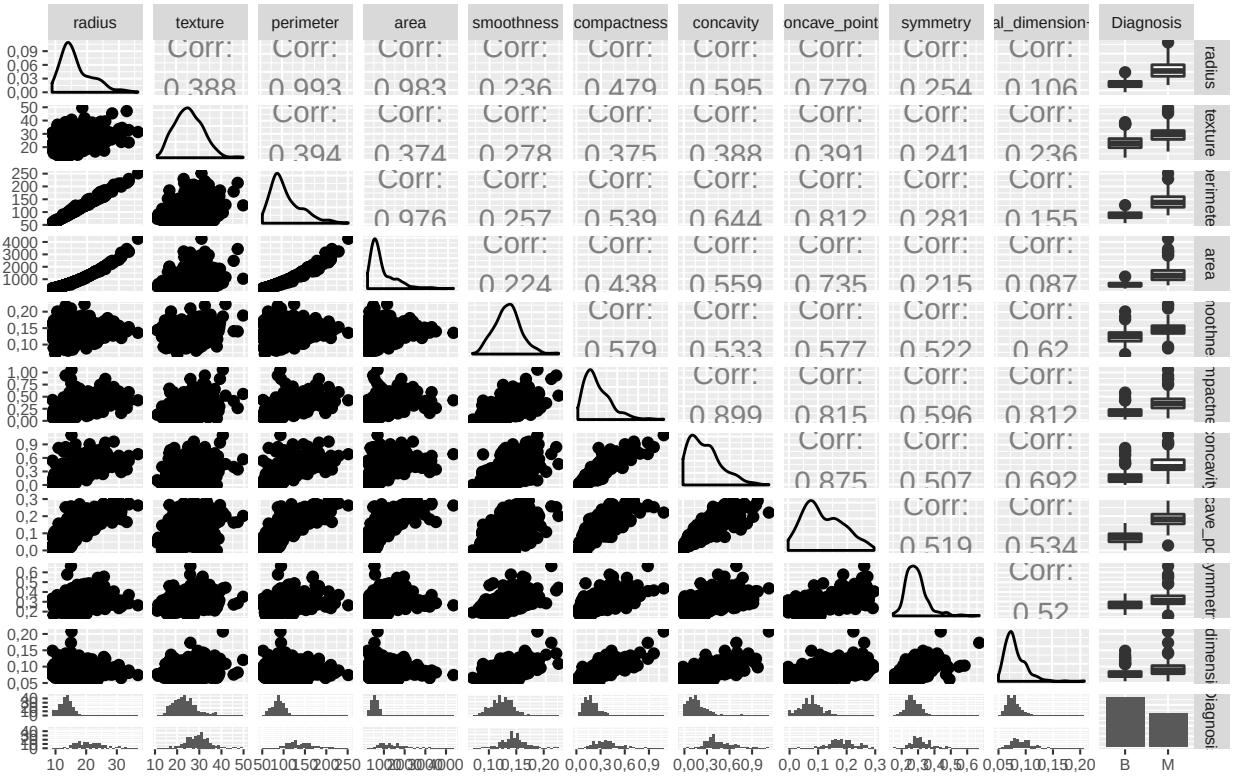
Correlações

É possível observar que há muitas variáveis bastante correlacionadas



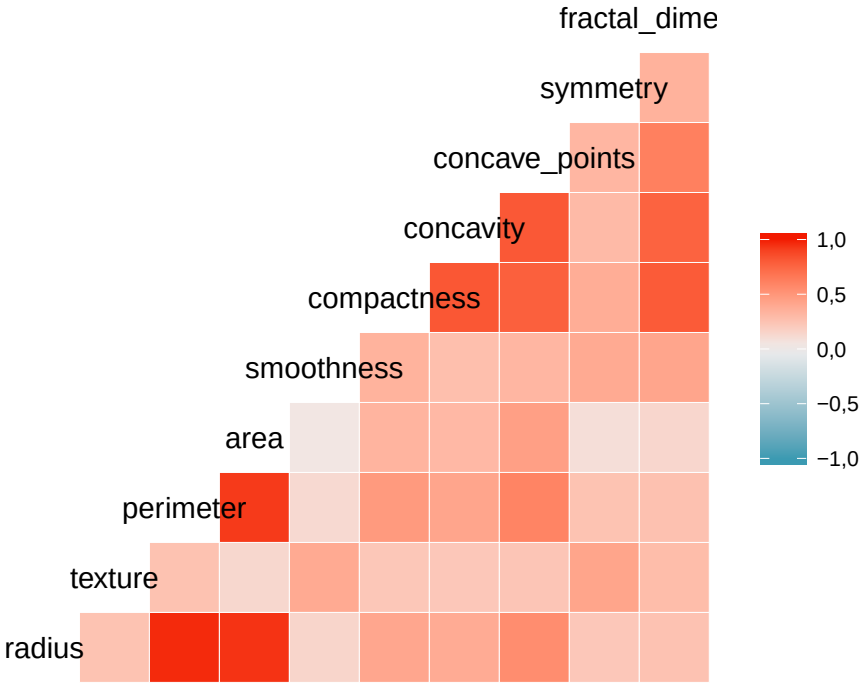
Resumo das distribuições

É possível observar que o valor da maioria das variáveis se mostra maior para os resultados malignos



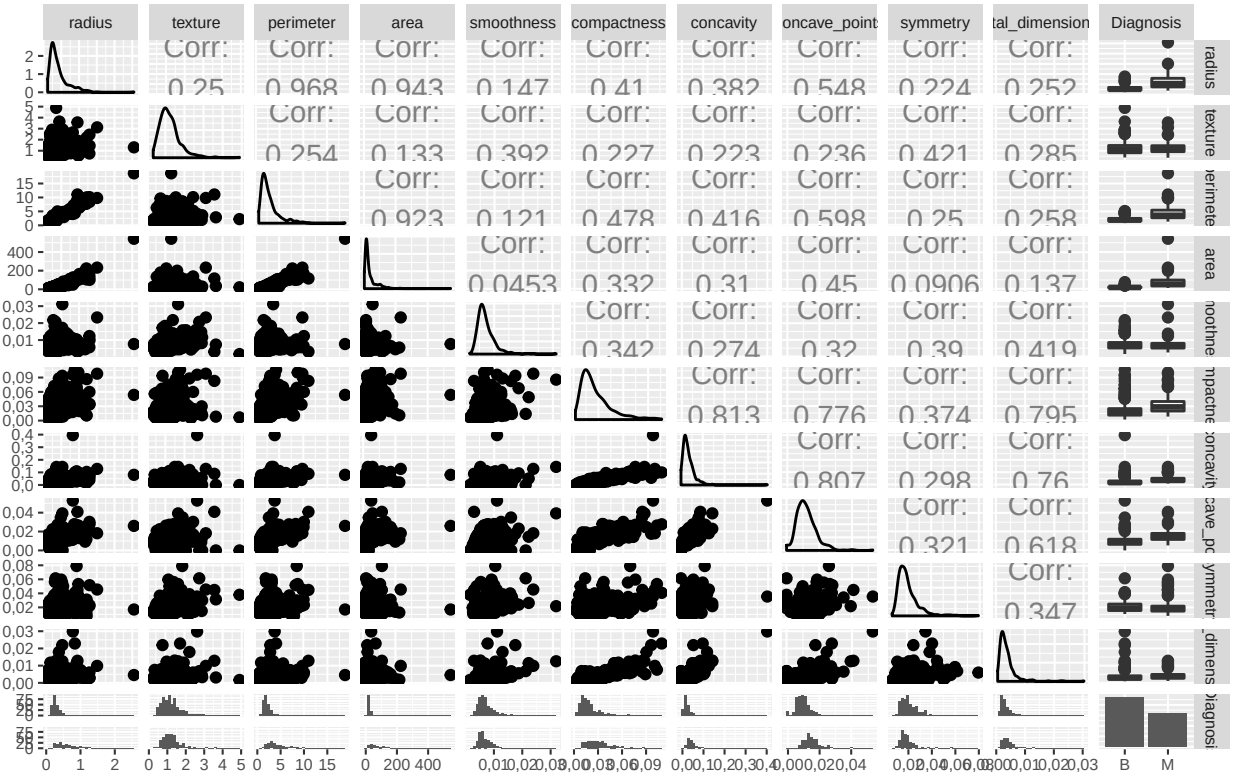
Correlações

É possível observar que há muitas variáveis bastante correlacionadas



Resumo das distribuições

É possível observar que o valor da maioria das variáveis se mostra maior para os resultados malignos



3 Modelos de classificação

O modelo é testado em esquema 4-Fold-validation repetido 5 vezes, para melhor avaliação da variância do modelo

3.1 Regressão logística

Em seguida é mostrado o resultado da regressão logística sem o tratamento de redução de dimensionalidade PCA.

```
## # A tibble: 20 x 2
##    Resample       n
##    <chr>      <int>
##  1 Fold1.Rep1   100
##  2 Fold1.Rep2   100
##  3 Fold1.Rep3    99
##  4 Fold1.Rep4   100
##  5 Fold1.Rep5   100
##  6 Fold2.Rep1    99
##  7 Fold2.Rep2    99
##  8 Fold2.Rep3   100
##  9 Fold2.Rep4   100
## 10 Fold2.Rep5    99
## 11 Fold3.Rep1   100
```

```
## 12 Fold3.Rep2   100
## 13 Fold3.Rep3   100
## 14 Fold3.Rep4    99
## 15 Fold3.Rep5    99
## 16 Fold4.Rep1    99
## 17 Fold4.Rep2    99
## 18 Fold4.Rep3    99
## 19 Fold4.Rep4    99
## 20 Fold4.Rep5   100
```

```
## [1] 398
```

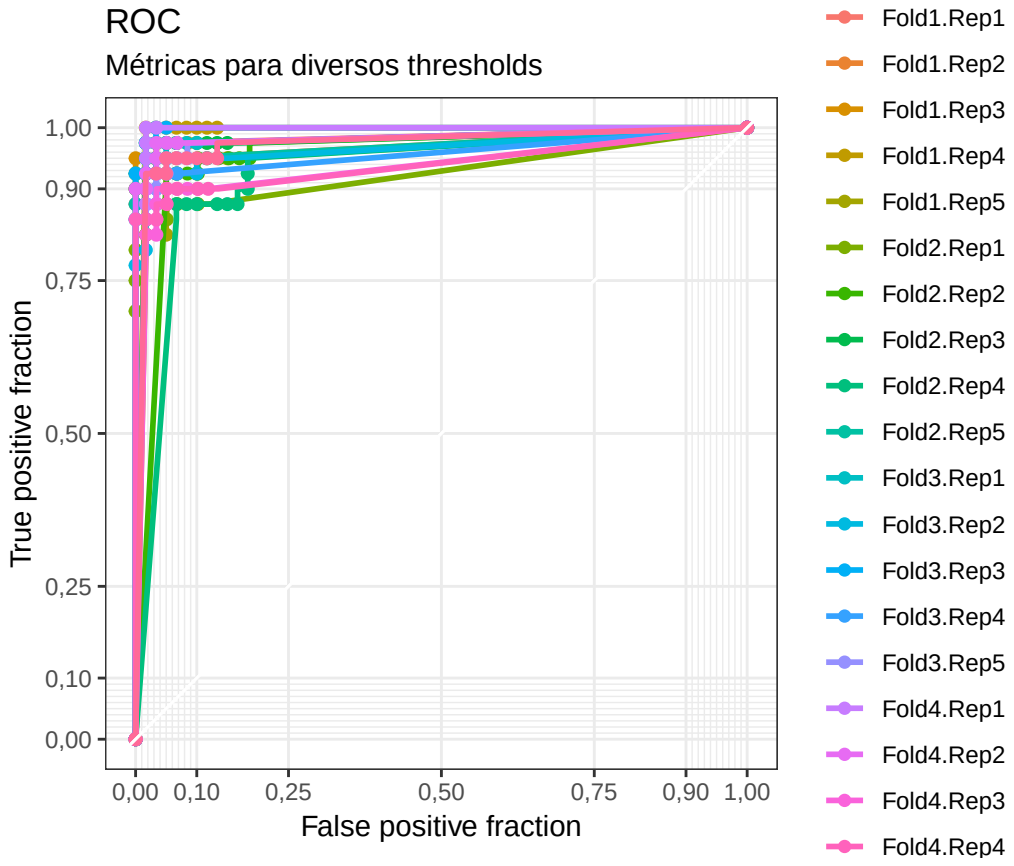


Tabela 1: Métricas para cada Fold

ROC	Sens	Spec	Resample
98.56%	96.67%	97.5%	Fold1.Rep1
92.80%	100.00%	75.0%	Fold2.Rep1
97.65%	95.00%	97.5%	Fold3.Rep1
99.75%	98.31%	90.0%	Fold4.Rep1
96.46%	96.67%	95.0%	Fold1.Rep2
95.44%	94.92%	92.5%	Fold2.Rep2
96.67%	98.33%	87.5%	Fold3.Rep2
98.43%	98.31%	92.5%	Fold4.Rep2
98.54%	100.00%	95.0%	Fold1.Rep3
97.71%	91.67%	97.5%	Fold2.Rep3
99.83%	98.33%	97.5%	Fold3.Rep3
93.43%	96.61%	90.0%	Fold4.Rep3

ROC	Sens	Spec	Resample
99.17%	95.00%	100.0%	Fold1.Rep4
92.71%	93.33%	87.5%	Fold2.Rep4
95.78%	98.31%	87.5%	Fold3.Rep4
94.11%	94.92%	87.5%	Fold4.Rep4
95.62%	95.00%	87.5%	Fold1.Rep5
97.29%	98.31%	95.0%	Fold2.Rep5
97.44%	96.61%	92.5%	Fold3.Rep5
97.35%	93.33%	95.0%	Fold4.Rep5

Vemos um resultado com viés e variância ruins

Métrica	Média	Desvio-padrão
ROC	96.74%	2.18%
Sens	96.48%	2.29%
Spec	92.00%	5.71%

O modelo final mostra que nenhum coeficiente tem significado estatístico. Isso pode ser um efeito colateral de um número muito grande de variáveis explicativas.

```
##
## Call:
## NULL
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1,100e-04  -2,100e-08  -2,100e-08   2,100e-08   1,102e-04
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -3,474e+01  6,286e+04  -0,001    1,000
## radius_media             -1,184e+03  3,203e+06   0,000    1,000
## texture_media            -5,003e+01  1,644e+05   0,000    1,000
## perimeter_media           1,270e+03  2,036e+06   0,001    1,000
## area_media                3,337e+01  1,873e+06   0,000    1,000
## smoothness_media          4,673e+01  7,219e+04   0,001    0,999
## compactness_media        -5,460e+01  1,006e+05  -0,001    1,000
## concavity_media           5,121e+01  2,869e+05   0,000    1,000
## concave_points_media     -6,286e+01  1,115e+05  -0,001    1,000
## symmetry_media           -1,427e+00  6,414e+04   0,000    1,000
## fractal_dimension_media -3,603e+01  1,234e+05   0,000    1,000
## radius_dv                -2,088e+00  5,645e+05   0,000    1,000
## texture_dv               -5,184e+01  5,308e+04  -0,001    0,999
## perimeter_dv              9,548e+00  8,960e+04   0,000    1,000
## area_dv                   1,108e+02  8,100e+05   0,000    1,000
## smoothness_dv             3,412e+01  7,158e+04   0,000    1,000
## compactness_dv            2,160e+01  3,107e+05   0,000    1,000
## concavity_dv              7,147e+01  1,436e+05   0,000    1,000
## concave_points_dv         2,962e+01  1,036e+05   0,000    1,000
## symmetry_dv              -3,747e+00  9,340e+04   0,000    1,000
## fractal_dimension_dv     -1,664e+02  4,099e+05   0,000    1,000
## radius_pior               6,436e+02  4,502e+05   0,001    0,999
```

```
## texture_pior             1,211e+02  1,915e+05   0,001    0,999
## perimeter_pior          -2,842e+02  4,701e+05  -0,001    1,000
## area_pior               -3,910e+02  5,980e+05  -0,001    0,999
## smoothness_pior         -3,273e+01  9,340e+04   0,000    1,000
## compactness_pior        -4,017e+01  3,389e+05   0,000    1,000
## concavity_pior          -4,827e+01  1,913e+05   0,000    1,000
## concave_points_pior      1,636e+01  1,268e+05   0,000    1,000
## symmetry_pior            1,096e+01  7,041e+04   0,000    1,000
## fractal_dimension_pior   1,263e+02  2,749e+05   0,000    1,000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5,3636e+02  on 397  degrees of freedom
## Residual deviance: 9,1816e-08  on 367  degrees of freedom
## AIC: 62
##
## Number of Fisher Scoring iterations: 25
```

Existem alguns pares de variáveis extremamente correlacionados. A chance de que seja necessário ter as duas variáveis é pequena.

```
## # A tibble: 435 x 3
##    coluna1          coluna2          correlacao
##    <chr>            <chr>                 <dbl>
##  1 perimeter_media radius_media          0.998
##  2 perimeter_pior  radius_pior           0.993
##  3 area_media      radius_media          0.988
##  4 area_media      perimeter_media       0.987
##  5 area_pior       radius_pior           0.983
##  6 area_pior       perimeter_pior        0.976
##  7 perimeter_media perimeter_pior        0.971
##  8 radius_media    radius_pior           0.971
##  9 perimeter_media radius_pior           0.970
## 10 perimeter_dv    radius_dv             0.968
## # ... with 425 more rows
```

Para cada par de variáveis mais correlacionadas x_i, x_j , vamos testar um modelo sem a variável x_i e outro sem a variável x_j. A variável que participar do modelo de pior resultado é marcada para exclusão.

par	correlacao	valor	formula	media	dp	maior_media
1	0,9976477	perimeter_media	Diagnosis ~ . -perimeter_media	0,9687945	0.015861537006685	0,9687945
1	0,9976477	radius_media	Diagnosis ~ . -radius_media	0,9661891	0.0198066625974758	0,9687945
2	0,9932071	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9728775	0.0192935944605359	0,9728775
2	0,9932071	radius_pior	Diagnosis ~ . -radius_pior	0,9665002	0.0182631901375293	0,9728775
3	0,9878539	area_media	Diagnosis ~ . -area_media	0,9703386	0.0223691266032003	0,9703386
3	0,9878539	radius_media	Diagnosis ~ . -radius_media	0,9616197	0.0159326238896229	0,9703386
4	0,9868384	perimeter_media	Diagnosis ~ . -perimeter_media	0,9695782	0.0172935018020441	0,9695782
4	0,9868384	area_media	Diagnosis ~ . -area_media	0,9685016	0.0150742005938626	0,9695782
5	0,9830428	radius_pior	Diagnosis ~ . -radius_pior	0,9678626	0.0193251435487687	0,9678626
5	0,9830428	area_pior	Diagnosis ~ . -area_pior	0,9649493	0.0304899596087297	0,9678626
6	0,9757201	area_pior	Diagnosis ~ . -area_pior	0,9743300	0.0188165954986588	0,9743300
6	0,9757201	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9731824	0.0167268052131208	0,9743300
7	0,9709805	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9691229	0.0231163416828619	0,9691229
7	0,9709805	perimeter_media	Diagnosis ~ . -perimeter_media	0,9638303	0.0176854994828509	0,9691229
8	0,9707329	radius_media	Diagnosis ~ . -radius_media	0,9749875	0.0161861778998064	0,9749875

par	correlacao	valor	formula	media	dp	maior_media
8	0,9707329	radius_pior	Diagnosis ~ . -radius_pior	0,9582218	0.0243816633997301	0,9749875
9	0,9702958	radius_pior	Diagnosis ~ . -radius_pior	0,9686651	0.0223980483759098	0,9686651
9	0,9702958	perimeter_media	Diagnosis ~ . -perimeter_media	0,9625989	0.0240027075504464	0,9686651
10	0,9682968	radius_dv	Diagnosis ~ . -radius_dv	0,9730856	0.0129995225521963	0,9730856
10	0,9682968	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9701900	0.0181047506707165	0,9730856
11	0,9678492	area_media	Diagnosis ~ . -area_media	0,9728430	0.0195917132498205	0,9728430
11	0,9678492	radius_pior	Diagnosis ~ . -radius_pior	0,9698840	0.0209599227127962	0,9728430
12	0,9654936	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9757777	0.0164244160998291	0,9757777
12	0,9654936	radius_media	Diagnosis ~ . -radius_media	0,9705320	0.0176880564264987	0,9757777
13	0,9633932	area_media	Diagnosis ~ . -area_media	0,9779248	0.0170338783528284	0,9779248
13	0,9633932	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9688127	0.0142271860390045	0,9779248
14	0,9622437	area_media	Diagnosis ~ . -area_media	0,9693639	0.0188759763233753	0,9693639
14	0,9622437	area_pior	Diagnosis ~ . -area_pior	0,9673065	0.0186354296106542	0,9693639

Modelo sem as variáveis altamente correlacionadas

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in verify_d(data$d): D not labeled 0/1, assuming B = 0 and M = 1!
```

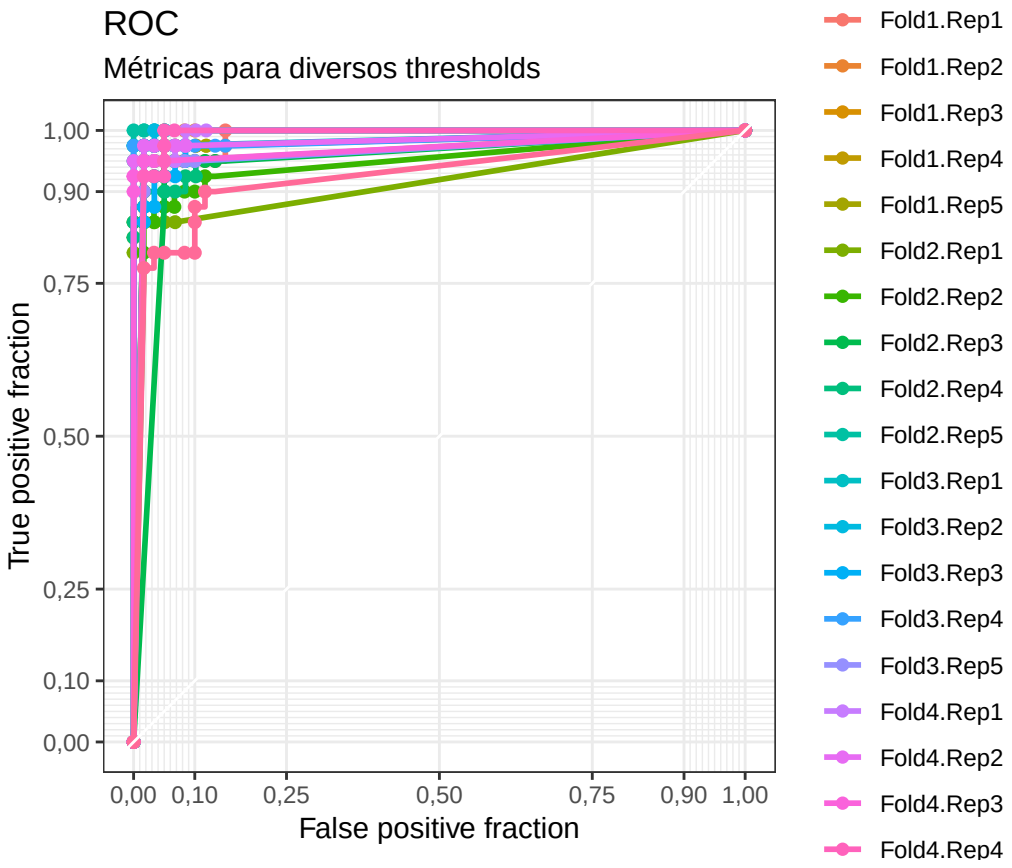


Tabela 4: Métricas para cada Fold

ROC	Sens	Spec	Resample
98.35%	93.3%	95.0%	Fold1.Rep1
91.99%	100.0%	85.0%	Fold2.Rep1
97.85%	98.3%	97.5%	Fold3.Rep1
98.81%	93.2%	95.0%	Fold4.Rep1
99.54%	96.7%	92.5%	Fold1.Rep2
94.44%	93.3%	90.0%	Fold2.Rep2
99.07%	94.9%	100.0%	Fold3.Rep2
98.60%	98.3%	97.5%	Fold4.Rep2
97.75%	93.2%	97.5%	Fold1.Rep3
94.56%	95.0%	95.0%	Fold2.Rep3
98.03%	96.6%	90.0%	Fold3.Rep3
97.33%	98.3%	92.5%	Fold4.Rep3
96.53%	96.6%	95.0%	Fold1.Rep4
96.74%	100.0%	87.5%	Fold2.Rep4
98.54%	95.0%	97.5%	Fold3.Rep4
98.85%	95.0%	92.5%	Fold4.Rep4
97.73%	98.3%	97.5%	Fold1.Rep5
100.00%	100.0%	100.0%	Fold2.Rep5
96.42%	96.7%	95.0%	Fold3.Rep5
92.54%	90.0%	82.5%	Fold4.Rep5

Métrica	Média	Desvio-padrão
ROC	97.18%	2.22%
Sens	96.14%	2.72%
Spec	93.75%	4.76%

Além de melhorar a performance (viés) do do modelo, a especificação sem as variáveis altamente correlacionadas mostra coeficientes estatisticamente significativos.

```
##
## Call:
## NULL
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1,610e-04  -2,100e-08  -2,100e-08   2,100e-08   1,864e-04
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  51,841  55526,086   0,001    0,999
## texture_media               -92,849  56789,999  -0,002    0,999
## smoothness_media             30,993  60344,602   0,001    1,000
## compactness_media            49,271  50929,692   0,001    0,999
## concavity_media             111,362  72953,436   0,002    0,999
## concave_points_media        -24,992  79591,527   0,000    1,000
## symmetry_media              -19,211  16334,518  -0,001    0,999
## fractal_dimension_media     -48,242  42351,834  -0,001    0,999
## radius_dv                   -31,777 273664,068   0,000    1,000
## texture_dv                  -83,272  69900,579  -0,001    0,999
## area_dv                     305,121 351679,537   0,001    0,999
## smoothness_dv                29,221  74089,183   0,000    1,000
## compactness_dv               -8,351 108661,752   0,000    1,000
## concavity_dv                112,198 137849,183   0,001    0,999
## concave_points_dv             6,470  75634,689   0,000    1,000
## symmetry_dv                  11,806  55746,193   0,000    1,000
## fractal_dimension_dv       -188,766 101632,215  -0,002    0,999
## texture_pior                192,341  55623,662   0,003    0,997
## smoothness_pior             -40,479  53662,881  -0,001    0,999
## compactness_pior            -80,115  96352,127  -0,001    0,999
## concavity_pior              -64,799  71922,353  -0,001    0,999
## concave_points_pior          27,270 120481,997   0,000    1,000
## symmetry_pior                20,565  13683,687   0,002    0,999
## fractal_dimension_pior      168,383  60418,607   0,003    0,998
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5,3636e+02  on 397  degrees of freedom
## Residual deviance: 2,0163e-07  on 374  degrees of freedom
## AIC: 48
##
## Number of Fisher Scoring iterations: 25
```

par	correlacao	valor	formula	media	dp	ma
1	0,9976477	perimeter_media	Diagnosis ~ . -perimeter_media	0,9734972	0.0172136020109249	

par	correlacao	valor	formula	media	dp	ma
1	0,9976477	radius_media	Diagnosis ~ . -radius_media	0,9708565	0.0152421074128869	
2	0,9932071	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9718305	0.0214076752332322	
2	0,9932071	radius_pior	Diagnosis ~ . -radius_pior	0,9686555	0.010701058604833	
3	0,9878539	radius_media	Diagnosis ~ . -radius_media	0,9697994	0.0208878733669911	
3	0,9878539	area_media	Diagnosis ~ . -area_media	0,9673679	0.0160554781173677	
4	0,9868384	perimeter_media	Diagnosis ~ . -perimeter_media	0,9725320	0.0189426450620923	
4	0,9868384	area_media	Diagnosis ~ . -area_media	0,9686469	0.0170371883394165	
5	0,9830428	area_pior	Diagnosis ~ . -area_pior	0,9673877	0.0209644160327419	
5	0,9830428	radius_pior	Diagnosis ~ . -radius_pior	0,9668533	0.0178347656393667	
6	0,9757201	area_pior	Diagnosis ~ . -area_pior	0,9738109	0.024562726889682	
6	0,9757201	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9670621	0.0232149709297019	
7	0,9709805	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9693307	0.0173165812232016	
7	0,9709805	perimeter_media	Diagnosis ~ . -perimeter_media	0,9670701	0.0113482008337006	
8	0,9707329	radius_pior	Diagnosis ~ . -radius_pior	0,9657235	0.0171978571966259	
8	0,9707329	radius_media	Diagnosis ~ . -radius_media	0,9644500	0.0251495305599341	
9	0,9702958	perimeter_media	Diagnosis ~ . -perimeter_media	0,9673430	0.0191200246668904	
9	0,9702958	radius_pior	Diagnosis ~ . -radius_pior	0,9652110	0.0190058903156366	
10	0,9682968	radius_dv	Diagnosis ~ . -radius_dv	0,9759034	0.0146301907000549	
10	0,9682968	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9645328	0.0190593802356607	
11	0,9678492	area_media	Diagnosis ~ . -area_media	0,9709126	0.0151401966285424	
11	0,9678492	radius_pior	Diagnosis ~ . -radius_pior	0,9687761	0.0230998361160171	
12	0,9654936	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9717675	0.0183621693447102	
12	0,9654936	radius_media	Diagnosis ~ . -radius_media	0,9714472	0.0163799400010854	
13	0,9633932	area_media	Diagnosis ~ . -area_media	0,9707874	0.0165670453666491	
13	0,9633932	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9694462	0.0138446630187798	
14	0,9622437	area_pior	Diagnosis ~ . -area_pior	0,9728724	0.0132969125336875	
14	0,9622437	area_media	Diagnosis ~ . -area_media	0,9721301	0.019905930721985	
15	0,9433532	radius_dv	Diagnosis ~ . -radius_dv	0,9788505	0.0169864208629686	
15	0,9433532	area_dv	Diagnosis ~ . -area_dv	0,9708923	0.0140566440948385	
16	0,9393057	radius_media	Diagnosis ~ . -radius_media	0,9701910	0.0185225913229853	
16	0,9393057	area_pior	Diagnosis ~ . -area_pior	0,9696939	0.0170530459782554	
17	0,9390764	perimeter_media	Diagnosis ~ . -perimeter_media	0,9668291	0.0153563013575635	
17	0,9390764	area_pior	Diagnosis ~ . -area_pior	0,9655069	0.0206583920890632	
18	0,9274525	concave_points_media	Diagnosis ~ . -concave_points_media	0,9725305	0.0137487743009756	
18	0,9274525	concavity_media	Diagnosis ~ . -concavity_media	0,9665056	0.0228788549741313	
19	0,9227395	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9733701	0.015055533674169	
19	0,9227395	area_dv	Diagnosis ~ . -area_dv	0,9663598	0.0203153743928032	
20	0,9131443	texture_media	Diagnosis ~ . -texture_media	0,9660825	0.0153960899879707	
20	0,9131443	texture_pior	Diagnosis ~ . -texture_pior	0,9611089	0.0157828262856292	
21	0,9119390	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9724380	0.0189868923290078	
21	0,9119390	concave_points_media	Diagnosis ~ . -concave_points_media	0,9671374	0.0234844485512181	

Modelo sem as variáveis altamente correlacionadas

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in verify_d(data$d): D not labeled 0/1, assuming B = 0 and M = 1!
```

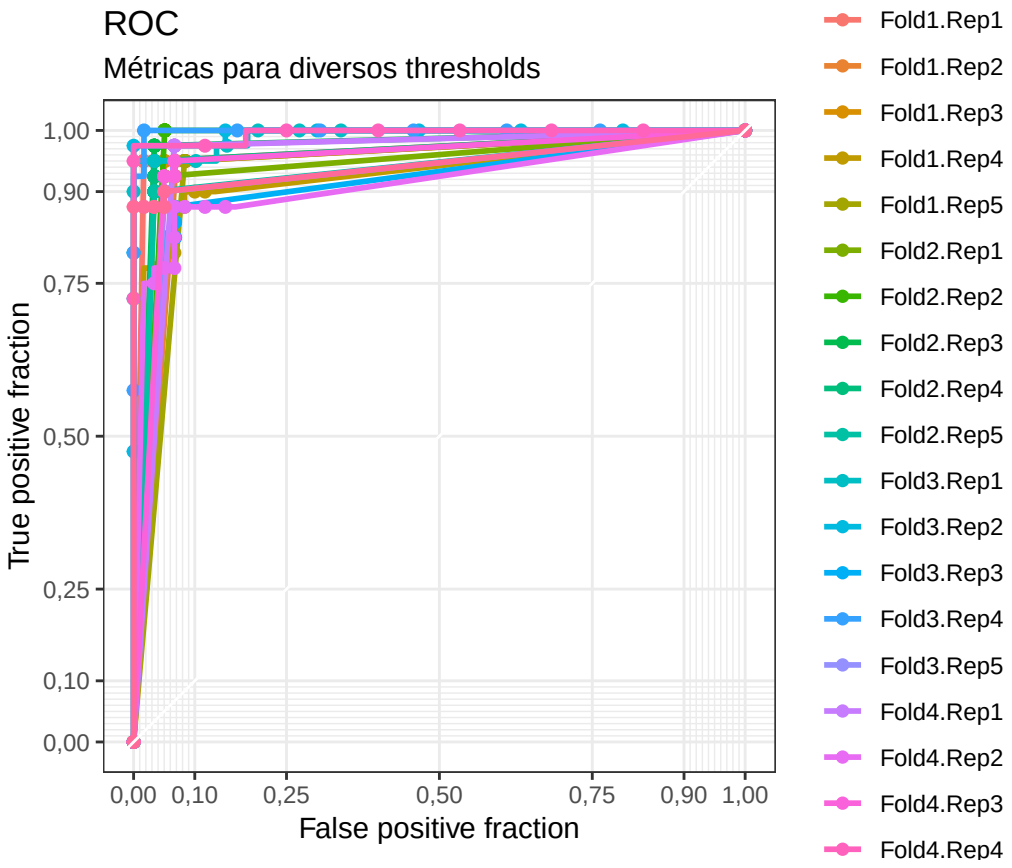


Tabela 7: Métricas para cada Fold

ROC	Sens	Spec	Resample
96.62%	98.33%	95.0%	Fold1.Rep1
94.45%	96.61%	92.5%	Fold2.Rep1
99.19%	96.61%	95.0%	Fold3.Rep1
95.42%	93.33%	97.5%	Fold4.Rep1
93.98%	91.53%	95.0%	Fold1.Rep2
97.46%	94.92%	100.0%	Fold2.Rep2
99.62%	100.00%	97.5%	Fold3.Rep2
91.33%	93.33%	77.5%	Fold4.Rep2
94.96%	95.00%	95.0%	Fold1.Rep3
97.06%	96.61%	97.5%	Fold2.Rep3
90.89%	93.22%	85.0%	Fold3.Rep3
94.85%	95.00%	92.5%	Fold4.Rep3
92.85%	93.33%	82.5%	Fold1.Rep4
95.76%	96.61%	92.5%	Fold2.Rep4
99.87%	94.92%	100.0%	Fold3.Rep4
99.54%	98.33%	97.5%	Fold4.Rep4
93.33%	91.67%	95.0%	Fold1.Rep5
93.31%	96.61%	90.0%	Fold2.Rep5
92.46%	94.92%	90.0%	Fold3.Rep5
94.62%	100.00%	87.5%	Fold4.Rep5

Métrica	Média	Desvio-padrão
ROC	95.38%	2.74%
Sens	95.54%	2.45%
Spec	92.75%	5.90%

Além de melhorar a performance (viés) do do modelo, a especificação sem as variáveis altamente correlacionadas mostra coeficientes estatisticamente significativos.

```
##
## Call:
## NULL
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1,78742  -0,00197  -0,00001   0,00000   2,31730
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1,4242     1,0191   1,398  0,16226
## texture_media              3,1554     1,5158   2,082  0,03737 *
## smoothness_media          -1,4593     2,3611  -0,618  0,53653
## compactness_media          8,0577     4,2749   1,885  0,05945 .
## symmetry_media            -2,1754     1,4355  -1,515  0,12967
## fractal_dimension_media   -3,1799     3,1041  -1,024  0,30564
## radius_dv                 12,8366     4,7473   2,704  0,00685 **
## texture_dv                -0,4879     1,5876  -0,307  0,75861
## smoothness_dv              4,1447     1,8438   2,248  0,02458 *
## compactness_dv             0,6246     4,0805   0,153  0,87834
## concavity_dv               1,8471     5,9651   0,310  0,75683
## concave_points_dv         -3,9414     2,6538  -1,485  0,13749
## symmetry_dv               -2,3566     1,7631  -1,337  0,18136
## fractal_dimension_dv      -8,8277     5,6377  -1,566  0,11739
## smoothness_pior           -2,5940     2,5511  -1,017  0,30923
## compactness_pior         -14,9572     5,7520  -2,600  0,00931 **
## concavity_pior             4,9620     4,3464   1,142  0,25360
## concave_points_pior       14,9956     6,1541   2,437  0,01482 *
## symmetry_pior              4,5300     1,8850   2,403  0,01625 *
## fractal_dimension_pior     9,2802     5,0497   1,838  0,06609 .
## ---
## Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 536,359  on 397  degrees of freedom
## Residual deviance:  27,813  on 378  degrees of freedom
## AIC: 67,813
##
## Number of Fisher Scoring iterations: 12
```

par	correlacao	valor	formula	media	dp	ma
1	0,9976477	radius_media	Diagnosis ~ . -radius_media	0,9656804	0.0196837238105445	
1	0,9976477	perimeter_media	Diagnosis ~ . -perimeter_media	0,9629015	0.0169531050679003	
2	0,9932071	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9741402	0.0197751475987448	

par	correlacao	valor	formula	media	dp	ma
2	0,9932071	radius_pior	Diagnosis ~ . -radius_pior	0,9661254	0.0161508201335915	
3	0,9878539	area_media	Diagnosis ~ . -area_media	0,9727857	0.0239127459869462	
3	0,9878539	radius_media	Diagnosis ~ . -radius_media	0,9674770	0.0200475806232386	
4	0,9868384	area_media	Diagnosis ~ . -area_media	0,9721472	0.0138085151865711	
4	0,9868384	perimeter_media	Diagnosis ~ . -perimeter_media	0,9656409	0.0189973547166543	
5	0,9830428	area_pior	Diagnosis ~ . -area_pior	0,9697918	0.0194065194368715	
5	0,9830428	radius_pior	Diagnosis ~ . -radius_pior	0,9612698	0.0211180081125001	
6	0,9757201	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9701559	0.0165000114063681	
6	0,9757201	area_pior	Diagnosis ~ . -area_pior	0,9673264	0.0240786563411224	
7	0,9709805	perimeter_media	Diagnosis ~ . -perimeter_media	0,9695019	0.0174232676031315	
7	0,9709805	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9692509	0.0205913553346683	
8	0,9707329	radius_media	Diagnosis ~ . -radius_media	0,9687209	0.0185143603425739	
8	0,9707329	radius_pior	Diagnosis ~ . -radius_pior	0,9654610	0.0188761680967341	
9	0,9702958	perimeter_media	Diagnosis ~ . -perimeter_media	0,9679550	0.0170579992493434	
9	0,9702958	radius_pior	Diagnosis ~ . -radius_pior	0,9650321	0.0206328928040911	
10	0,9682968	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9699578	0.0178963005064784	
10	0,9682968	radius_dv	Diagnosis ~ . -radius_dv	0,9693978	0.0174092815538055	
11	0,9678492	area_media	Diagnosis ~ . -area_media	0,9694320	0.017292921537797	
11	0,9678492	radius_pior	Diagnosis ~ . -radius_pior	0,9597814	0.0200543841423814	
12	0,9654936	radius_media	Diagnosis ~ . -radius_media	0,9699970	0.0239373648343764	
12	0,9654936	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9688570	0.0176499904874119	
13	0,9633932	area_media	Diagnosis ~ . -area_media	0,9734161	0.0171204770700965	
13	0,9633932	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9675614	0.0170696915402322	
14	0,9622437	area_pior	Diagnosis ~ . -area_pior	0,9745230	0.0186132731319589	
14	0,9622437	area_media	Diagnosis ~ . -area_media	0,9685360	0.0217657518861475	
15	0,9433532	radius_dv	Diagnosis ~ . -radius_dv	0,9719440	0.01577383993826	
15	0,9433532	area_dv	Diagnosis ~ . -area_dv	0,9704608	0.0154161651170168	
16	0,9393057	radius_media	Diagnosis ~ . -radius_media	0,9699435	0.0192194487823599	
16	0,9393057	area_pior	Diagnosis ~ . -area_pior	0,9677300	0.02278577539635	
17	0,9390764	perimeter_media	Diagnosis ~ . -perimeter_media	0,9722413	0.0151434782933562	
17	0,9390764	area_pior	Diagnosis ~ . -area_pior	0,9688579	0.017310508210516	
18	0,9274525	concave_points_media	Diagnosis ~ . -concave_points_media	0,9711751	0.0186302600864215	
18	0,9274525	concavity_media	Diagnosis ~ . -concavity_media	0,9702943	0.0163680064218037	
19	0,9227395	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9710637	0.0229973093830904	
19	0,9227395	area_dv	Diagnosis ~ . -area_dv	0,9635118	0.0195578229623846	
20	0,9131443	texture_media	Diagnosis ~ . -texture_media	0,9729357	0.0172141974452366	
20	0,9131443	texture_pior	Diagnosis ~ . -texture_pior	0,9682498	0.0177686919161855	
21	0,9119390	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9752059	0.0152214535515547	
21	0,9119390	concave_points_media	Diagnosis ~ . -concave_points_media	0,9718150	0.0144670449894209	
22	0,8991618	concavity_media	Diagnosis ~ . -concavity_media	0,9719675	0.0181549735762831	
22	0,8991618	compactness_media	Diagnosis ~ . -compactness_media	0,9668280	0.017987444104707	
23	0,8989726	compactness_pior	Diagnosis ~ . -compactness_pior	0,9678365	0.0149176282966229	
23	0,8989726	concavity_pior	Diagnosis ~ . -concavity_pior	0,9662850	0.0184570951723958	
24	0,8921750	concavity_pior	Diagnosis ~ . -concavity_pior	0,9743257	0.0144024922887449	
24	0,8921750	concavity_media	Diagnosis ~ . -concavity_media	0,9715576	0.0194847882822869	
25	0,8754867	concavity_pior	Diagnosis ~ . -concavity_pior	0,9765692	0.0124808020369208	
25	0,8754867	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9740067	0.0162284859438587	
26	0,8653202	compactness_pior	Diagnosis ~ . -compactness_pior	0,9669799	0.0181013527943277	
26	0,8653202	compactness_media	Diagnosis ~ . -compactness_media	0,9627265	0.0236936866558619	
27	0,8639394	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9725044	0.0143942661036636	
27	0,8639394	concavity_media	Diagnosis ~ . -concavity_media	0,9713911	0.0146938190612955	
28	0,8573737	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9677542	0.0202040818259524	

par	correlacao	valor	formula	media	dp	ma
28	0,8573737	concave_points_media	Diagnosis ~ . -concave_points_media	0,9660049	0.0138825140042051	
29	0,8548703	area_pior	Diagnosis ~ . -area_pior	0,9756931	0.0198511044141216	
29	0,8548703	area_dv	Diagnosis ~ . -area_dv	0,9704880	0.0228114989678553	
30	0,8529464	concave_points_media	Diagnosis ~ . -concave_points_media	0,9715720	0.0171923321263773	
30	0,8529464	perimeter_media	Diagnosis ~ . -perimeter_media	0,9715055	0.017895346131641	

Modelo sem as variáveis altamente correlacionadas

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in verify_d(data$d): D not labeled 0/1, assuming B = 0 and M = 1!
```

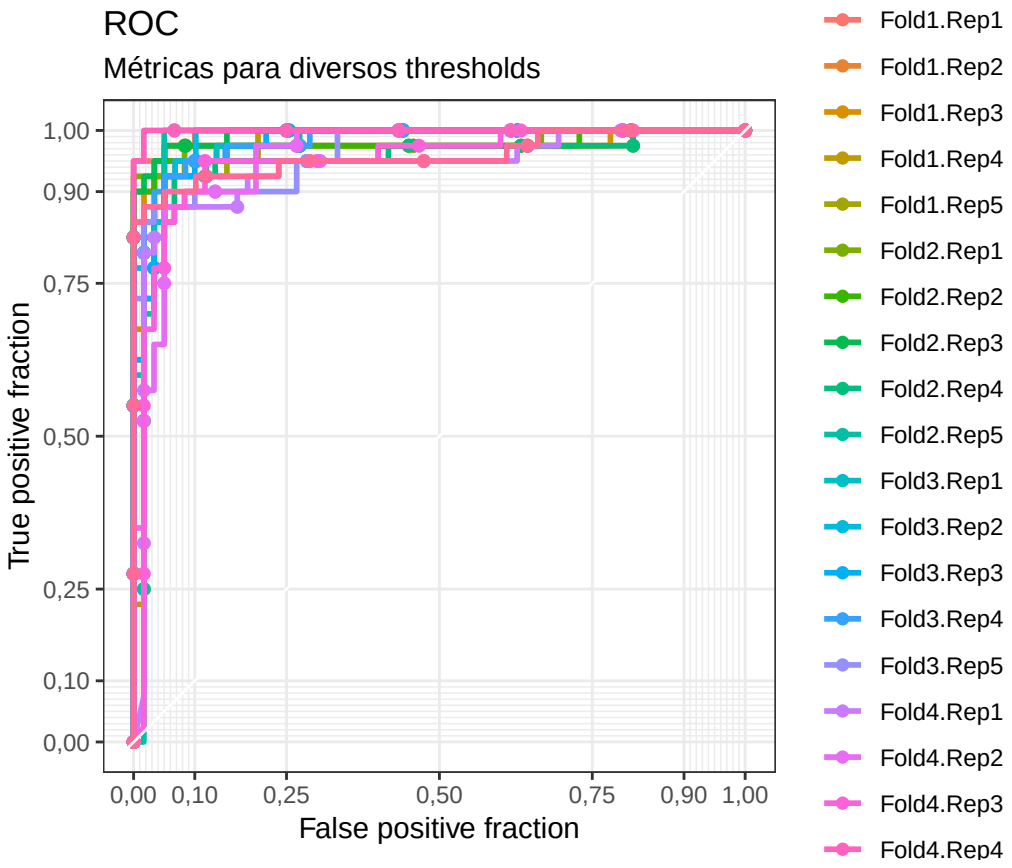


Tabela 10: Métricas para cada Fold

ROC	Sens	Spec	Resample
99.62%	93.3%	97.5%	Fold1.Rep1
97.92%	96.7%	92.5%	Fold2.Rep1
97.88%	94.9%	90.0%	Fold3.Rep1
93.58%	83.1%	87.5%	Fold4.Rep1
96.75%	100.0%	90.0%	Fold1.Rep2
97.67%	91.5%	95.0%	Fold2.Rep2
98.81%	94.9%	92.5%	Fold3.Rep2
95.88%	91.7%	90.0%	Fold4.Rep2
97.33%	94.9%	95.0%	Fold1.Rep3
99.36%	93.2%	97.5%	Fold2.Rep3
97.83%	93.3%	92.5%	Fold3.Rep3
94.73%	93.3%	87.5%	Fold4.Rep3
97.29%	83.1%	97.5%	Fold1.Rep4
96.08%	96.7%	85.0%	Fold2.Rep4
98.35%	100.0%	82.5%	Fold3.Rep4
99.92%	98.3%	95.0%	Fold4.Rep4
97.80%	86.4%	92.5%	Fold1.Rep5
97.62%	95.0%	95.0%	Fold2.Rep5
96.46%	96.7%	87.5%	Fold3.Rep5
95.81%	98.3%	87.5%	Fold4.Rep5

Métrica	Média	Desvio-padrão
ROC	97.33%	1.60%
Sens	93.77%	4.83%
Spec	91.50%	4.32%

Além de melhorar a performance (viés) do do modelo, a especificação sem as variáveis altamente correlacionadas mostra coeficientes estatisticamente significativos.

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2,1256  -0,1099  -0,0093   0,0223   2,8864
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -0,88616    0,41353  -2,143 0,032121 *
## texture_media            1,96245    0,50762   3,866 0,000111 ***
## smoothness_media         1,45837    0,90509   1,611 0,107112
## symmetry_media          -0,50730    0,63830  -0,795 0,426747
## fractal_dimension_media -3,31562    1,10874  -2,990 0,002786 **
## texture_dv              -0,75308    0,52800  -1,426 0,153783
## perimeter_dv             4,36000    0,94928   4,593 4,37e-06 ***
## smoothness_dv            0,09036    0,59580   0,152 0,879449
## compactness_dv          -2,94110    1,34643  -2,184 0,028935 *
## concavity_dv             2,39678    0,94190   2,545 0,010940 *
## concave_points_dv        2,42531    0,91104   2,662 0,007765 **
## symmetry_dv             -0,53647    0,75567  -0,710 0,477750
## fractal_dimension_dv    -2,91654    1,61274  -1,808 0,070539 .
## smoothness_pior          0,18098    1,03259   0,175 0,860872
## compactness_pior         1,09090    1,69749   0,643 0,520450
## symmetry_pior            1,51482    0,80088   1,891 0,058566 .
## fractal_dimension_pior   3,47051    1,72592   2,011 0,044345 *
## ---
## Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 536,359  on 397  degrees of freedom
## Residual deviance:  86,755  on 381  degrees of freedom
## AIC: 120,76
##
## Number of Fisher Scoring iterations: 8
```

par	correlacao	valor	formula	media	dp	r
1	0,9976477	radius_media	Diagnosis ~ . -radius_media	0,9744576	0.0177806405464611	
1	0,9976477	perimeter_media	Diagnosis ~ . -perimeter_media	0,9650207	0.0191447780730162	
2	0,9932071	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9659244	0.0155828757577654	
2	0,9932071	radius_pior	Diagnosis ~ . -radius_pior	0,9625925	0.0210184466267934	
3	0,9878539	area_media	Diagnosis ~ . -area_media	0,9760035	0.0199278155591721	
3	0,9878539	radius_media	Diagnosis ~ . -radius_media	0,9729605	0.0144829455488574	

par	correlacao	valor	formula	media	dp	r
4	0,9868384	area_media	Diagnosis ~ . -area_media	0,9692449	0.0195394197326731	
4	0,9868384	perimeter_media	Diagnosis ~ . -perimeter_media	0,9669587	0.0198828959601354	
5	0,9830428	area_pior	Diagnosis ~ . -area_pior	0,9711301	0.0163934638699416	
5	0,9830428	radius_pior	Diagnosis ~ . -radius_pior	0,9671029	0.0155792595009685	
6	0,9757201	area_pior	Diagnosis ~ . -area_pior	0,9695037	0.0165152130429613	
6	0,9757201	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9621792	0.0182416996229161	
7	0,9709805	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9656933	0.0209419356965009	
7	0,9709805	perimeter_media	Diagnosis ~ . -perimeter_media	0,9631623	0.0172035927063955	
8	0,9707329	radius_pior	Diagnosis ~ . -radius_pior	0,9703379	0.0218256512827673	
8	0,9707329	radius_media	Diagnosis ~ . -radius_media	0,9700673	0.0183022190904494	
9	0,9702958	perimeter_media	Diagnosis ~ . -perimeter_media	0,9738282	0.0131095211467089	
9	0,9702958	radius_pior	Diagnosis ~ . -radius_pior	0,9685874	0.0141808557692272	
10	0,9682968	radius_dv	Diagnosis ~ . -radius_dv	0,9771545	0.0159065554130128	
10	0,9682968	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9724375	0.0192629830146043	
11	0,9678492	area_media	Diagnosis ~ . -area_media	0,9804762	0.0106995831244093	
11	0,9678492	radius_pior	Diagnosis ~ . -radius_pior	0,9689398	0.0186849198845418	
12	0,9654936	radius_media	Diagnosis ~ . -radius_media	0,9682812	0.0194490339378322	
12	0,9654936	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9678678	0.0184476487747241	
13	0,9633932	area_media	Diagnosis ~ . -area_media	0,9718199	0.0177186486491759	
13	0,9633932	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9700960	0.0228253031155132	
14	0,9622437	area_pior	Diagnosis ~ . -area_pior	0,9743355	0.0146609439265437	
14	0,9622437	area_media	Diagnosis ~ . -area_media	0,9725703	0.0118167483131032	
15	0,9433532	radius_dv	Diagnosis ~ . -radius_dv	0,9769308	0.0145612145956105	
15	0,9433532	area_dv	Diagnosis ~ . -area_dv	0,9682145	0.0177916944279079	
16	0,9393057	radius_media	Diagnosis ~ . -radius_media	0,9732627	0.0152296543210738	
16	0,9393057	area_pior	Diagnosis ~ . -area_pior	0,9707405	0.018093666563294	
17	0,9390764	perimeter_media	Diagnosis ~ . -perimeter_media	0,9769225	0.016882169626045	
17	0,9390764	area_pior	Diagnosis ~ . -area_pior	0,9647398	0.0257917476416661	
18	0,9274525	concavity_media	Diagnosis ~ . -concavity_media	0,9723416	0.0176889253731125	
18	0,9274525	concave_points_media	Diagnosis ~ . -concave_points_media	0,9718409	0.0177714560716759	
19	0,9227395	area_dv	Diagnosis ~ . -area_dv	0,9702608	0.0160496108226827	
19	0,9227395	perimeter_dv	Diagnosis ~ . -perimeter_dv	0,9667108	0.0208198810655101	
20	0,9131443	texture_media	Diagnosis ~ . -texture_media	0,9703845	0.0173641514239985	
20	0,9131443	texture_pior	Diagnosis ~ . -texture_pior	0,9628275	0.0221185874260911	
21	0,9119390	concave_points_media	Diagnosis ~ . -concave_points_media	0,9742440	0.0132553302952349	
21	0,9119390	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9714477	0.0180198711628273	
22	0,8991618	compactness_media	Diagnosis ~ . -compactness_media	0,9759737	0.0178033078037204	
22	0,8991618	concavity_media	Diagnosis ~ . -concavity_media	0,9667703	0.0127637231873238	
23	0,8989726	concavity_pior	Diagnosis ~ . -concavity_pior	0,9701096	0.0209676300969604	
23	0,8989726	compactness_pior	Diagnosis ~ . -compactness_pior	0,9684292	0.0199121201555838	
24	0,8921750	concavity_media	Diagnosis ~ . -concavity_media	0,9749562	0.0149763438386622	
24	0,8921750	concavity_pior	Diagnosis ~ . -concavity_pior	0,9688413	0.0193967753592814	
25	0,8754867	concavity_pior	Diagnosis ~ . -concavity_pior	0,9717961	0.0148781844895651	
25	0,8754867	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9679590	0.0245419504161157	
26	0,8653202	compactness_pior	Diagnosis ~ . -compactness_pior	0,9724479	0.0169228790298638	
26	0,8653202	compactness_media	Diagnosis ~ . -compactness_media	0,9636308	0.0164661898350165	
27	0,8639394	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9741144	0.0216702946783901	
27	0,8639394	concavity_media	Diagnosis ~ . -concavity_media	0,9661637	0.0244619363229964	
28	0,8573737	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9693176	0.0172592007487256	
28	0,8573737	concave_points_media	Diagnosis ~ . -concave_points_media	0,9638741	0.0182367193906641	
29	0,8548703	area_pior	Diagnosis ~ . -area_pior	0,9655388	0.0188075602426709	
29	0,8548703	area_dv	Diagnosis ~ . -area_dv	0,9655312	0.0237017169693413	

par	correlacao	valor	formula	media	dp	r
30	0,8529464	concave_points_media	Diagnosis ~ . -concave_points_media	0,9663450	0.0163373978458207	
30	0,8529464	perimeter_media	Diagnosis ~ . -perimeter_media	0,9647632	0.021058752038555	
31	0,8460189	compactness_media	Diagnosis ~ . -compactness_media	0,9694793	0.0185885590837443	
31	0,8460189	concave_points_media	Diagnosis ~ . -concave_points_media	0,9669387	0.016503979729018	
32	0,8341235	compactness_media	Diagnosis ~ . -compactness_media	0,9689449	0.0206969494307046	
32	0,8341235	concavity_pior	Diagnosis ~ . -concavity_pior	0,9688264	0.0158817929202074	
33	0,8291507	concave_points_media	Diagnosis ~ . -concave_points_media	0,9768563	0.0185786566366331	
33	0,8291507	radius_pior	Diagnosis ~ . -radius_pior	0,9656351	0.0235081127007357	
34	0,8275557	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9690150	0.0177434729028319	
34	0,8275557	compactness_media	Diagnosis ~ . -compactness_media	0,9685554	0.0227140567142163	
35	0,8242218	area_media	Diagnosis ~ . -area_media	0,9726379	0.0187589804912354	
35	0,8242218	concave_points_media	Diagnosis ~ . -concave_points_media	0,9693257	0.0148784743683964	
36	0,8232562	concave_points_media	Diagnosis ~ . -concave_points_media	0,9729131	0.0188278396951155	
36	0,8232562	radius_media	Diagnosis ~ . -radius_media	0,9666886	0.0163693117574263	
37	0,8191072	smoothness_media	Diagnosis ~ . -smoothness_media	0,9693559	0.0175623971073474	
37	0,8191072	smoothness_pior	Diagnosis ~ . -smoothness_pior	0,9641840	0.0200101173807836	
38	0,8146701	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9709230	0.0194255739863998	
38	0,8146701	compactness_pior	Diagnosis ~ . -compactness_pior	0,9698063	0.0161916751123966	
39	0,8126923	concavity_dv	Diagnosis ~ . -concavity_dv	0,9760607	0.00980536189917312	
39	0,8126923	compactness_dv	Diagnosis ~ . -compactness_dv	0,9728102	0.0156318417822682	
40	0,8124397	compactness_pior	Diagnosis ~ . -compactness_pior	0,9698090	0.0171532361126486	
40	0,8124397	fractal_dimension_pior	Diagnosis ~ . -fractal_dimension_pior	0,9619101	0.0129809020881887	
41	0,8121101	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9702318	0.0211117987192119	
41	0,8121101	concave_points_pior	Diagnosis ~ . -concave_points_pior	0,9684322	0.016875741948159	
42	0,8073326	concave_points_dv	Diagnosis ~ . -concave_points_dv	0,9715369	0.0190197608227658	
42	0,8073326	concavity_dv	Diagnosis ~ . -concavity_dv	0,9658425	0.012783162573286	
43	0,8063195	area_media	Diagnosis ~ . -area_media	0,9704901	0.0177290269329615	
43	0,8063195	area_dv	Diagnosis ~ . -area_dv	0,9640946	0.0182462066377905	
44	0,8026914	perimeter_pior	Diagnosis ~ . -perimeter_pior	0,9671907	0.0202674390461332	
44	0,8026914	area_dv	Diagnosis ~ . -area_dv	0,9651880	0.0233446986935229	
45	0,8026389	concave_points_media	Diagnosis ~ . -concave_points_media	0,9717654	0.0162079641269687	
45	0,8026389	area_pior	Diagnosis ~ . -area_pior	0,9691049	0.0207151658302786	

Modelo sem as variáveis altamente correlacionadas

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in verify_d(data$d): D not labeled 0/1, assuming B = 0 and M = 1!
```

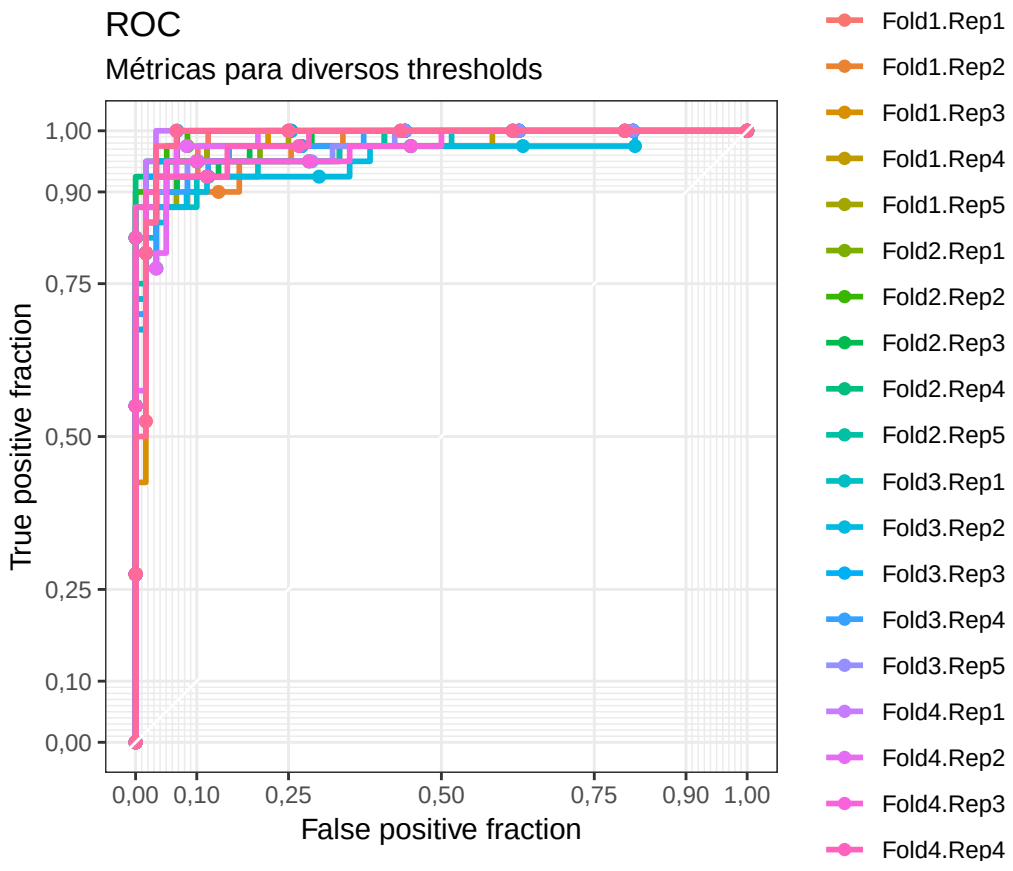


Tabela 13: Métricas para cada Fold

ROC	Sens	Spec	Resample
98.90%	98.31%	82.5%	Fold1.Rep1
97.67%	91.53%	95.0%	Fold2.Rep1
96.88%	93.33%	87.5%	Fold3.Rep1
99.58%	98.33%	95.0%	Fold4.Rep1
97.29%	93.22%	87.5%	Fold1.Rep2
99.49%	98.31%	92.5%	Fold2.Rep2
95.25%	93.33%	87.5%	Fold3.Rep2

ROC	Sens	Spec	Resample
98.12%	91.67%	97.5%	Fold4.Rep2
98.12%	91.67%	95.0%	Fold1.Rep3
98.22%	98.31%	87.5%	Fold2.Rep3
99.15%	98.31%	90.0%	Fold3.Rep3
97.46%	95.00%	95.0%	Fold4.Rep3
97.21%	96.67%	92.5%	Fold1.Rep4
99.62%	94.92%	95.0%	Fold2.Rep4
97.58%	94.92%	90.0%	Fold3.Rep4
98.42%	98.33%	87.5%	Fold4.Rep4
98.25%	93.33%	90.0%	Fold1.Rep5
98.22%	98.31%	87.5%	Fold2.Rep5
97.71%	96.61%	87.5%	Fold3.Rep5
98.83%	93.33%	97.5%	Fold4.Rep5

Métrica	Média	Desvio-padrão
ROC	98.10%	1.05%
Sens	95.39%	2.59%
Spec	91.00%	4.17%

Além de melhorar a performance (viés) do do modelo, a especificação sem as variáveis altamente correlacionadas mostra coeficientes estatisticamente significativos.

```
##
## Call:
## NULL
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2,49600  -0,13869  -0,01398   0,03561   2,80909
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -0,7581     0,3756  -2,018  0,04357 *
## texture_media             2,0555     0,4503   4,565 5,01e-06 ***
## smoothness_media          1,4760     0,5722   2,580  0,00989 **
## symmetry_media           -1,3648     0,5816  -2,347  0,01895 *
## fractal_dimension_media  -1,0368     0,7281  -1,424  0,15444
## radius_dv                 4,2544     0,8518   4,995 5,89e-07 ***
## texture_dv               -1,0789     0,4917  -2,194  0,02821 *
## smoothness_dv            -0,3296     0,5039  -0,654  0,51309
## concave_points_dv         2,6715     0,6454   4,139 3,49e-05 ***
## symmetry_dv              -2,4454     0,5393  -4,535 5,77e-06 ***
## fractal_dimension_dv     -0,9546     0,6360  -1,501  0,13340
## symmetry_pior             3,9688     0,7111   5,581 2,39e-08 ***
## ---
## Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 536,36  on 397  degrees of freedom
```

```
## Residual deviance: 100,87  on 386  degrees of freedom
## AIC: 124,87
##
## Number of Fisher Scoring iterations: 8
```