Classification Problem

Xinzhu Li

Introduction

We used the Breast Cancer dataset to solve the classification problem. The analysis consists of 4 parts:

1: logistic regression;

2: decision tree;

3: random forest;

4: SVM

These 4 methods were applied to the Breast Cancer dataset to compare their differences and gain a better understanding of each approach.

Logistic Regression

First, we'll load the data and check if there are any missing values in the dataset. After that, we'll transform the data into the proper format. Then, we'll split the dataset into a training set and a test set, with a ratio of 70% for training and 30% for testing.

```r
library("mlbench")
data("BreastCancer")
head(BreastCancer)
```

```
       Id Cl.thickness Cell.size Cell.shape Marg.adhesion Epith.c.size
1 1000025            5         1          1             1            2
2 1002945            5         4          4             5            7
3 1015425            3         1          1             1            2
4 1016277            6         8          8             1            3
5 1017023            4         1          1             3            2
```

```
6 1017122            8        10         10             8            7
  Bare.nuclei Bl.cromatin Normal.nucleoli Mitoses     Class
1           1           3               1       1    benign
2          10           3               2       1    benign
3           2           3               1       1    benign
4           4           3               7       1    benign
5           1           3               1       1    benign
6          10           9               7       1 malignant
```

```
names(BreastCancer)
```

```
 [1] "Id"              "Cl.thickness"    "Cell.size"       "Cell.shape"
 [5] "Marg.adhesion"   "Epith.c.size"    "Bare.nuclei"     "Bl.cromatin"
 [9] "Normal.nucleoli" "Mitoses"         "Class"
```

```
str(BreastCancer)
```

```
'data.frame':   699 obs. of  11 variables:
 $ Id             : chr  "1000025" "1002945" "1015425" "1016277" ...
 $ Cl.thickness   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 5 5 3 6 4 8 1 2 2 4 ...
 $ Cell.size      : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 4 1 8 1 10 1 1 1 2 ...
 $ Cell.shape     : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 4 1 8 1 10 1 2 1 1 ...
 $ Marg.adhesion  : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 5 1 1 3 8 1 1 1 1 ...
 $ Epith.c.size   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 2 7 2 3 2 7 2 2 2 2 ...
 $ Bare.nuclei    : Factor w/ 10 levels "1","2","3","4",..: 1 10 2 4 1 10 10 1 1 1 ...
 $ Bl.cromatin    : Factor w/ 10 levels "1","2","3","4",..: 3 3 3 3 3 9 3 3 1 2 ...
 $ Normal.nucleoli: Factor w/ 10 levels "1","2","3","4",..: 1 2 1 7 1 7 1 1 1 1 ...
 $ Mitoses        : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 5 1 ...
 $ Class          : Factor w/ 2 levels "benign","malignant": 1 1 1 1 1 2 1 1 1 1 ...
```

```
df <- BreastCancer[-1]
df
```

```
    Cl.thickness Cell.size Cell.shape Marg.adhesion Epith.c.size Bare.nuclei
1              5         1          1             1            2           1
2              5         4          4             5            7          10
3              3         1          1             1            2           2
4              6         8          8             1            3           4
5              4         1          1             3            2           1
6              8        10         10             8            7          10
```

7	1	1	1	1	2	10
8	2	1	2	1	2	1
9	2	1	1	1	2	1
10	4	2	1	1	2	1
11	1	1	1	1	1	1
12	2	1	1	1	2	1
13	5	3	3	3	2	3
14	1	1	1	1	2	3
15	8	7	5	10	7	9
16	7	4	6	4	6	1
17	4	1	1	1	2	1
18	4	1	1	1	2	1
19	10	7	7	6	4	10
20	6	1	1	1	2	1
21	7	3	2	10	5	10
22	10	5	5	3	6	7
23	3	1	1	1	2	1
24	8	4	5	1	2	<NA>
25	1	1	1	1	2	1
26	5	2	3	4	2	7
27	3	2	1	1	1	1
28	5	1	1	1	2	1
29	2	1	1	1	2	1
30	1	1	3	1	2	1
31	3	1	1	1	1	1
32	2	1	1	1	2	1
33	10	7	7	3	8	5
34	2	1	1	2	2	1
35	3	1	2	1	2	1
36	2	1	1	1	2	1
37	10	10	10	8	6	1
38	6	2	1	1	1	1
39	5	4	4	9	2	10
40	2	5	3	3	6	7
41	6	6	6	9	6	<NA>
42	10	4	3	1	3	3
43	6	10	10	2	8	10
44	5	6	5	6	10	1
45	10	10	10	4	8	1
46	1	1	1	1	2	1
47	3	7	7	4	4	9
48	1	1	1	1	2	1
49	4	1	1	3	2	1

50	7	8	7	2	4	8
51	9	5	8	1	2	3
52	5	3	3	4	2	4
53	10	3	6	2	3	5
54	5	5	5	8	10	8
55	10	5	5	6	8	8
56	10	6	6	3	4	5
57	8	10	10	1	3	6
58	8	2	4	1	5	1
59	5	2	3	1	6	10
60	9	5	5	2	2	2
61	5	3	5	5	3	3
62	1	1	1	1	2	2
63	9	10	10	1	10	8
64	6	3	4	1	5	2
65	1	1	1	1	2	1
66	10	4	2	1	3	2
67	4	1	1	1	2	1
68	5	3	4	1	8	10
69	8	3	8	3	4	9
70	1	1	1	1	2	1
71	5	1	3	1	2	1
72	6	10	2	8	10	2
73	1	3	3	2	2	1
74	9	4	5	10	6	10
75	10	6	4	1	3	4
76	1	1	2	1	2	2
77	1	1	4	1	2	1
78	5	3	1	2	2	1
79	3	1	1	1	2	3
80	2	1	1	1	3	1
81	2	2	2	1	1	1
82	4	1	1	2	2	1
83	5	2	1	1	2	1
84	3	1	1	1	2	2
85	3	5	7	8	8	9
86	5	10	6	1	10	4
87	3	3	6	4	5	8
88	3	6	6	6	5	10
89	4	1	1	1	2	1
90	2	1	1	2	3	1
91	1	1	1	1	2	1
92	3	1	1	2	2	1

93	4	1	1	1	2	1
94	1	1	1	1	2	1
95	2	1	1	1	2	1
96	1	1	1	1	2	1
97	2	1	1	2	2	1
98	5	1	1	1	2	1
99	9	6	9	2	10	6
100	7	5	6	10	5	10
101	10	3	5	1	10	5
102	2	3	4	4	2	5
103	4	1	2	1	2	1
104	8	2	3	1	6	3
105	10	10	10	10	10	1
106	7	3	4	4	3	3
107	10	10	10	8	2	10
108	1	6	8	10	8	10
109	1	1	1	1	2	1
110	6	5	4	4	3	9
111	1	3	1	2	2	2
112	8	6	4	3	5	9
113	10	3	3	10	2	10
114	10	10	10	3	10	8
115	3	3	2	1	2	3
116	1	1	1	1	2	5
117	8	3	3	1	2	2
118	4	5	5	10	4	10
119	1	1	1	1	4	3
120	3	2	1	1	2	2
121	1	1	2	2	2	1
122	4	2	1	1	2	2
123	10	10	10	2	10	10
124	5	3	5	1	8	10
125	5	4	6	7	9	7
126	1	1	1	1	2	1
127	7	5	3	7	4	10
128	3	1	1	1	2	1
129	8	3	5	4	5	10
130	1	1	1	1	10	1
131	5	1	3	1	2	1
132	2	1	1	1	2	1
133	5	10	8	10	8	10
134	3	1	1	1	2	1
135	3	1	1	1	3	1

```
136            5         1           1             1            2           2
137            4         1           1             1            2           1
138            3         1           1             1            2           1
139            4         1           2             1            2           1
140            1         1           1             1            1        <NA>
141            3         1           1             1            2           1
142            2         1           1             1            2           1
143            9         5           5             4            4           5
144            1         1           1             1            2           5
145            2         1           1             1            2           1
146            1         1           3             1            2        <NA>
147            3         4           5             2            6           8
148            1         1           1             1            3           2
149            3         1           1             3            8           1
150            8         8           7             4           10          10
151            1         1           1             1            1           1
152            7         2           4             1            6          10
153           10        10           8             6            4           5
154            4         1           1             1            2           3
155            1         1           1             1            2           1
156            5         5           5             6            3          10
157            1         2           2             1            2           1
158            2         1           1             1            2           1
159            1         1           2             1            3        <NA>
160            9         9          10             3            6          10
161           10         7           7             4            5          10
162            4         1           1             1            2           1
163            3         1           1             1            2           1
164            1         1           1             2            1           3
165            5         1           1             1            2        <NA>
166            4         1           1             1            2           2
167            5         6           7             8            8          10
168           10         8          10            10            6           1
169            3         1           1             1            2           1
170            1         1           1             2            1           1
171            3         1           1             1            2           1
172            1         1           1             1            2           1
173            1         1           1             1            2           1
174            6        10          10            10            8          10
175            8         6           5             4            3          10
176            5         8           7             7           10          10
177            2         1           1             1            2           1
178            5        10          10             3            8           1
```

179	4	1	1	1	2	1
180	5	3	3	3	6	10
181	1	1	1	1	1	1
182	1	1	1	1	2	1
183	6	1	1	1	2	1
184	5	8	8	8	5	10
185	8	7	6	4	4	10
186	2	1	1	1	1	1
187	1	5	8	6	5	8
188	10	5	6	10	6	10
189	5	8	4	10	5	8
190	1	2	3	1	2	1
191	10	10	10	8	6	8
192	7	5	10	10	10	10
193	5	1	1	1	2	1
194	1	1	1	1	2	1
195	3	1	1	1	2	1
196	4	1	1	1	2	1
197	8	4	4	5	4	7
198	5	1	1	4	2	1
199	1	1	1	1	2	1
200	3	1	1	1	2	1
201	9	7	7	5	5	10
202	10	8	8	4	10	10
203	1	1	1	1	2	1
204	5	1	1	1	2	1
205	1	1	1	1	2	1
206	5	10	10	9	6	10
207	10	10	9	3	7	5
208	1	1	1	1	1	1
209	1	1	1	1	1	1
210	5	1	1	1	1	1
211	8	10	10	10	5	10
212	8	10	8	8	4	8
213	1	1	1	1	2	1
214	10	10	10	10	7	10
215	10	10	10	10	3	10
216	8	7	8	7	5	5
217	1	1	1	1	2	1
218	1	1	1	1	2	1
219	6	10	7	7	6	4
220	6	1	3	1	2	1
221	1	1	1	2	2	1

```
222      10     6     4     3    10    10
223       4     1     1     3     1     5
224       7     5     6     3     3     8
225      10     5     5     6     3    10
226       1     1     1     1     2     1
227      10     5     7     4     4    10
228       8     9     9     5     3     5
229       1     1     1     1     1     1
230      10    10    10     3    10    10
231       7     4     7     4     3     7
232       6     8     7     5     6     8
233       8     4     6     3     3     1
234      10     4     5     5     5    10
235       3     3     2     1     3     1
236       3     1     4     1     2  <NA>
237      10     8     8     2     8    10
238       9     8     8     5     6     2
239       8    10    10     8     6     9
240      10     4     3     2     3    10
241       5     1     3     3     2     2
242       3     1     1     3     1     1
243       2     1     1     1     2     1
244       1     1     1     1     2     5
245       1     1     1     1     2     1
246       5     1     1     2     2     2
247       8    10    10     8     5    10
248       8     4     4     1     2     9
249       4     1     1     1     2     1
250       3     1     1     1     2  <NA>
251       1     2     2     1     2     1
252      10     4     4    10     2    10
253       6     3     3     5     3    10
254       6    10    10     2     8    10
255       9    10    10     1    10     8
256       5     6     6     2     4    10
257       3     1     1     1     2     1
258       3     1     1     1     2     1
259       3     1     1     1     2     1
260       5     7     7     1     5     8
261      10     5     8    10     3    10
262       5    10    10     6    10    10
263       8     8     9     4     5    10
264      10     4     4    10     6    10
```

265	7	9	4	10	10	3
266	5	1	4	1	2	1
267	10	10	6	3	3	10
268	3	3	5	2	3	10
269	10	8	8	2	3	4
270	1	1	1	1	2	1
271	8	4	7	1	3	10
272	5	1	1	1	2	1
273	3	3	5	2	3	10
274	7	2	4	1	3	4
275	3	1	1	1	2	1
276	3	1	3	1	2	<NA>
277	3	1	1	1	2	1
278	1	1	1	1	2	1
279	1	1	1	1	2	1
280	10	5	7	3	3	7
281	3	1	1	1	2	1
282	2	1	1	2	2	1
283	1	4	3	10	4	10
284	10	4	6	1	2	10
285	7	4	5	10	2	10
286	8	10	10	10	8	10
287	10	10	10	10	10	10
288	3	1	1	1	3	1
289	6	1	3	1	4	5
290	5	6	6	8	6	10
291	1	1	1	1	2	1
292	1	1	1	1	2	1
293	8	8	8	1	2	<NA>
294	10	4	4	6	2	10
295	1	1	1	1	2	<NA>
296	5	5	7	8	6	10
297	5	3	4	3	4	5
298	5	4	3	1	2	<NA>
299	8	2	1	1	5	1
300	9	1	2	6	4	10
301	8	4	10	5	4	4
302	1	1	1	1	2	1
303	10	10	10	7	9	10
304	1	1	1	1	2	1
305	8	3	4	9	3	10
306	10	8	4	4	4	10
307	1	1	1	1	2	1

308	1	1	1	1	2	1
309	7	8	7	6	4	3
310	3	1	1	1	2	5
311	2	1	1	1	3	1
312	1	1	1	1	2	1
313	8	6	4	10	10	1
314	1	1	1	1	2	1
315	1	1	1	1	1	1
316	4	6	5	6	7	<NA>
317	5	5	5	2	5	10
318	6	8	7	8	6	8
319	1	1	1	1	5	1
320	4	4	4	4	6	5
321	7	6	3	2	5	10
322	3	1	1	1	2	<NA>
323	3	1	1	1	2	1
324	5	4	6	10	2	10
325	1	1	1	1	2	1
326	3	2	2	1	2	1
327	10	1	1	1	2	10
328	1	1	1	1	2	1
329	8	10	3	2	6	4
330	10	4	6	4	5	10
331	10	4	7	2	2	8
332	5	1	1	1	2	1
333	5	2	2	2	2	1
334	5	4	6	6	4	10
335	8	6	7	3	3	10
336	1	1	1	1	2	1
337	6	5	5	8	4	10
338	1	1	1	1	2	1
339	1	1	1	1	1	1
340	8	5	5	5	2	10
341	10	3	3	1	2	10
342	1	1	1	1	2	1
343	2	1	1	1	2	1
344	1	1	1	1	2	1
345	7	6	4	8	10	10
346	1	1	1	1	2	1
347	5	2	2	2	3	1
348	1	1	1	1	1	1
349	3	4	4	10	5	1
350	4	2	3	5	3	8

351	5	1	1	3	2	1
352	2	1	1	1	2	1
353	3	4	5	3	7	3
354	2	7	10	10	7	10
355	1	1	1	1	2	1
356	4	1	1	1	3	1
357	5	3	3	1	3	3
358	8	10	10	7	10	10
359	8	10	5	3	8	4
360	10	3	5	4	3	7
361	6	10	10	10	10	10
362	3	10	3	10	6	10
363	3	2	2	1	4	3
364	4	4	4	2	2	3
365	2	1	1	1	2	1
366	2	1	1	1	2	1
367	6	10	10	10	8	10
368	5	8	8	10	5	10
369	1	1	3	1	2	1
370	1	1	3	1	1	1
371	4	3	2	1	3	1
372	1	1	3	1	2	1
373	4	1	2	1	2	1
374	5	1	1	2	2	1
375	3	1	2	1	2	1
376	1	1	1	1	2	1
377	1	1	1	1	2	1
378	1	1	1	1	1	1
379	3	1	1	4	3	1
380	5	3	4	1	4	1
381	1	1	1	1	2	1
382	10	6	3	6	4	10
383	3	2	2	2	2	1
384	2	1	1	1	2	1
385	2	1	1	1	2	1
386	3	3	2	2	3	1
387	7	6	6	3	2	10
388	5	3	3	2	3	1
389	2	1	1	1	2	1
390	5	1	1	1	3	2
391	1	1	1	2	2	1
392	10	8	7	4	3	10
393	3	1	1	1	2	1

```
394              1         1          1             1           1           1
395              1         2          3             1           2           1
396              3         1          1             1           2           1
397              3         1          1             1           2           1
398              4         1          1             1           2           1
399              3         2          1             1           2           1
400              1         2          3             1           2           1
401              3        10          8             7           6           9
402              3         1          1             1           2           1
403              5         3          3             1           2           1
404              3         1          1             1           2           4
405              1         2          1             3           2           1
406              1         1          1             1           2           1
407              4         2          2             1           2           1
408              1         1          1             1           2           1
409              2         3          2             2           2           2
410              3         1          2             1           2           1
411              1         1          1             1           2           1
412              1         1          1             1           1        <NA>
413             10        10         10             6           8           4
414              5         1          2             1           2           1
415              8         5          6             2           3          10
416              3         3          2             6           3           3
417              8         7          8             5          10          10
418              1         1          1             1           2           1
419              5         2          2             2           2           2
420              2         3          1             1           5           1
421              3         2          2             3           2           3
422             10        10         10             7          10          10
423              4         3          3             1           2           1
424              5         1          3             1           2           1
425              3         1          1             1           2           1
426              9        10         10            10          10          10
427              5         3          6             1           2           1
428              8         7          8             2           4           2
429              1         1          1             1           2           1
430              2         1          1             1           2           1
431              1         3          1             1           2           1
432              5         1          1             3           4           1
433              5         1          1             1           2           1
434              3         2          2             3           2           1
435              6         9          7             5           5           8
436             10         8         10             1           3          10
```

437	10	10	10	1	6	1
438	4	1	1	1	2	1
439	4	1	3	3	2	1
440	5	1	1	1	2	1
441	10	4	3	10	4	10
442	5	2	2	4	2	4
443	1	1	1	3	2	3
444	1	1	1	1	2	2
445	5	1	1	6	3	1
446	2	1	1	1	2	1
447	1	1	1	1	2	1
448	5	1	1	1	2	1
449	1	1	1	1	1	1
450	5	7	9	8	6	10
451	4	1	1	3	1	1
452	5	1	1	1	2	1
453	3	1	1	3	2	1
454	4	5	5	8	6	10
455	2	3	1	1	3	1
456	10	2	2	1	2	6
457	10	6	5	8	5	10
458	8	8	9	6	6	3
459	5	1	2	1	2	1
460	5	1	3	1	2	1
461	5	1	1	3	2	1
462	3	1	1	1	2	5
463	6	1	1	3	2	1
464	4	1	1	1	2	1
465	4	1	1	1	2	1
466	10	9	8	7	6	4
467	10	6	6	2	4	10
468	6	6	6	5	4	10
469	4	1	1	1	2	1
470	1	1	2	1	2	1
471	3	1	1	1	1	1
472	6	1	1	3	2	1
473	6	1	1	1	1	1
474	4	1	1	1	2	1
475	5	1	1	1	2	1
476	3	1	1	1	2	1
477	4	1	2	1	2	1
478	4	1	1	1	2	1
479	5	2	1	1	2	1

480	4	8	7	10	4	10
481	5	1	1	1	1	1
482	5	3	2	4	2	1
483	9	10	10	10	10	5
484	8	7	8	5	5	10
485	5	1	2	1	2	1
486	1	1	1	3	1	3
487	3	1	1	1	1	1
488	10	10	10	10	6	10
489	3	6	4	10	3	3
490	6	3	2	1	3	4
491	1	1	1	1	2	1
492	5	8	9	4	3	10
493	4	1	1	1	1	1
494	5	10	10	10	6	10
495	5	1	2	10	4	5
496	3	1	1	1	1	1
497	1	1	1	1	1	1
498	4	2	1	1	2	1
499	4	1	1	1	2	1
500	4	1	1	1	2	1
501	6	1	1	1	2	1
502	4	1	1	1	2	1
503	4	1	1	2	2	1
504	4	1	1	1	2	1
505	1	1	1	1	2	1
506	3	3	1	1	2	1
507	8	10	10	10	7	5
508	1	1	1	1	2	4
509	5	1	1	1	2	1
510	2	1	1	1	2	1
511	1	1	1	1	2	1
512	5	1	1	1	2	1
513	5	1	1	1	2	1
514	3	1	1	1	1	1
515	6	6	7	10	3	10
516	4	10	4	7	3	10
517	1	1	1	1	1	1
518	1	1	1	1	1	1
519	3	1	2	2	2	1
520	4	7	8	3	4	10
521	1	1	1	1	3	1
522	4	1	1	1	3	1

523	10	4	5	4	3	5
524	7	5	6	10	4	10
525	3	1	1	1	2	1
526	3	1	1	2	2	1
527	4	1	1	1	2	1
528	4	1	1	1	2	1
529	6	1	3	2	2	1
530	4	1	1	1	1	1
531	7	4	4	3	4	10
532	4	2	2	1	2	1
533	1	1	1	1	1	1
534	3	1	1	1	2	1
535	2	1	1	1	2	1
536	1	1	3	2	2	1
537	5	1	1	1	2	1
538	5	1	2	1	2	1
539	4	1	1	1	2	1
540	6	1	1	1	2	1
541	5	1	1	1	2	2
542	3	1	1	1	2	1
543	5	3	1	1	2	1
544	4	1	1	1	2	1
545	2	1	3	2	2	1
546	5	1	1	1	2	1
547	6	10	10	10	4	10
548	2	1	1	1	1	1
549	3	1	1	1	1	1
550	7	8	3	7	4	5
551	3	1	1	1	2	1
552	1	1	1	1	2	1
553	3	2	2	2	2	1
554	4	4	2	1	2	5
555	3	1	1	1	2	1
556	4	3	1	1	2	1
557	5	2	2	2	1	1
558	5	1	1	3	2	1
559	2	1	1	1	2	1
560	5	1	1	1	2	1
561	5	1	1	1	2	1
562	5	1	1	1	2	1
563	1	1	1	1	2	1
564	3	1	1	1	2	1
565	4	1	1	1	2	1

566	5	7	10	10	5	10
567	3	1	2	1	2	1
568	4	1	1	1	2	3
569	8	4	4	1	6	10
570	10	10	8	10	6	5
571	8	10	4	4	8	10
572	7	6	10	5	3	10
573	3	1	1	1	2	1
574	1	1	1	1	2	1
575	10	9	7	3	4	2
576	5	1	2	1	2	1
577	5	1	1	1	2	1
578	1	1	1	1	2	1
579	1	1	1	1	2	1
580	1	1	1	1	2	1
581	5	1	2	1	2	1
582	5	7	10	6	5	10
583	6	10	5	5	4	10
584	3	1	1	1	2	1
585	5	1	1	6	3	1
586	1	1	1	1	2	1
587	8	10	10	10	6	10
588	5	1	1	1	2	1
589	9	8	8	9	6	3
590	5	1	1	1	2	1
591	4	10	8	5	4	1
592	2	5	7	6	4	10
593	10	3	4	5	3	10
594	5	1	2	1	2	1
595	4	8	6	3	4	10
596	5	1	1	1	2	1
597	4	1	2	1	2	1
598	5	1	3	1	2	1
599	3	1	1	1	2	1
600	5	2	4	1	1	1
601	3	1	1	1	2	1
602	1	1	1	1	1	1
603	4	1	1	1	2	1
604	5	4	6	8	4	1
605	5	3	2	8	5	10
606	10	5	10	3	5	8
607	4	1	1	2	2	1
608	1	1	1	1	2	1

609	5	10	10	10	10	10
610	5	1	1	1	2	1
611	10	4	3	10	3	10
612	5	10	10	10	5	2
613	8	10	10	10	6	10
614	2	3	1	1	2	1
615	2	1	1	1	1	1
616	4	1	3	1	2	1
617	3	1	1	1	2	1
618	1	1	1	1	1	<NA>
619	4	1	1	1	2	1
620	5	1	1	1	2	1
621	3	1	1	1	2	1
622	6	3	3	3	3	2
623	7	1	2	3	2	1
624	1	1	1	1	2	1
625	5	1	1	2	1	1
626	3	1	3	1	3	4
627	4	6	6	5	7	6
628	2	1	1	1	2	5
629	2	1	1	1	2	1
630	4	1	1	1	2	1
631	6	2	3	1	2	1
632	5	1	1	1	2	1
633	1	1	1	1	2	1
634	8	7	4	4	5	3
635	3	1	1	1	2	1
636	3	1	4	1	2	1
637	10	10	7	8	7	1
638	4	2	4	3	2	2
639	4	1	1	1	2	1
640	5	1	1	3	2	1
641	4	1	1	3	2	1
642	3	1	1	1	2	1
643	3	1	1	1	2	1
644	1	1	1	1	2	1
645	2	1	1	1	2	1
646	3	1	1	1	2	1
647	1	2	2	1	2	1
648	1	1	1	3	2	1
649	5	10	10	10	10	2
650	3	1	1	1	2	1
651	3	1	1	2	3	4

652	1	2	1	3	2	1
653	5	1	1	1	2	1
654	4	1	1	1	2	1
655	3	1	1	1	2	1
656	3	1	1	1	2	1
657	5	1	1	1	2	1
658	5	4	5	1	8	1
659	7	8	8	7	3	10
660	1	1	1	1	2	1
661	1	1	1	1	2	1
662	4	1	1	1	2	1
663	1	1	3	1	2	1
664	1	1	3	1	2	1
665	3	1	1	3	2	1
666	1	1	1	1	2	1
667	5	2	2	2	2	1
668	3	1	1	1	2	1
669	5	7	4	1	6	1
670	5	10	10	8	5	5
671	3	10	7	8	5	8
672	3	2	1	2	2	1
673	2	1	1	1	2	1
674	5	3	2	1	3	1
675	1	1	1	1	2	1
676	4	1	4	1	2	1
677	1	1	2	1	2	1
678	5	1	1	1	2	1
679	1	1	1	1	2	1
680	2	1	1	1	2	1
681	10	10	10	10	5	10
682	5	10	10	10	4	10
683	5	1	1	1	2	1
684	1	1	1	1	2	1
685	1	1	1	1	2	1
686	1	1	1	1	2	1
687	1	1	1	1	2	1
688	3	1	1	1	2	1
689	4	1	1	1	2	1
690	1	1	1	1	2	1
691	1	1	1	3	2	1
692	5	10	10	5	4	5
693	3	1	1	1	2	1
694	3	1	1	1	2	1

```
695            3         1          1             1            3           2
696            2         1          1             1            2           1
697            5        10         10             3            7           3
698            4         8          6             4            3           4
699            4         8          8             5            4           5
    Bl.cromatin Normal.nucleoli Mitoses     Class
1             3               1       1    benign
2             3               2       1    benign
3             3               1       1    benign
4             3               7       1    benign
5             3               1       1    benign
6             9               7       1 malignant
7             3               1       1    benign
8             3               1       1    benign
9             1               1       5    benign
10            2               1       1    benign
11            3               1       1    benign
12            2               1       1    benign
13            4               4       1 malignant
14            3               1       1    benign
15            5               5       4 malignant
16            4               3       1 malignant
17            2               1       1    benign
18            3               1       1    benign
19            4               1       2 malignant
20            3               1       1    benign
21            5               4       4 malignant
22            7              10       1 malignant
23            2               1       1    benign
24            7               3       1 malignant
25            3               1       1    benign
26            3               6       1 malignant
27            2               1       1    benign
28            2               1       1    benign
29            2               1       1    benign
30            1               1       1    benign
31            2               1       1    benign
32            3               1       1    benign
33            7               4       3 malignant
34            3               1       1    benign
35            2               1       1    benign
36            2               1       1    benign
37            8               9       1 malignant
```

```
38            7               1       1    benign
39            5               6       1 malignant
40            7               5       1 malignant
41            7               8       1    benign
42            6               5       2 malignant
43            7               3       3 malignant
44            3               1       1 malignant
45            8              10       1 malignant
46            2               1       2    benign
47            4               8       1 malignant
48            2               1       1    benign
49            3               1       1    benign
50            3               8       2 malignant
51            2               1       5 malignant
52            3               4       1 malignant
53            4              10       2 malignant
54            7               3       7 malignant
55            7               1       1 malignant
56            3               6       1 malignant
57            3               9       1 malignant
58            5               4       4 malignant
59            5               1       1 malignant
60            5               1       1 malignant
61            4              10       1 malignant
62            2               1       1    benign
63            3               3       1 malignant
64            3               9       1 malignant
65            2               1       1    benign
66            4               3      10 malignant
67            3               1       1    benign
68            4               9       1 malignant
69            8               9       8 malignant
70            3               2       1    benign
71            2               1       1    benign
72            7               8      10 malignant
73            7               2       1    benign
74            4               8       1 malignant
75            3               2       3 malignant
76            4               2       1    benign
77            2               1       1    benign
78            2               1       1    benign
79            3               1       1    benign
80            2               1       1    benign
```

```
81            7               1       1    benign
82            2               1       1    benign
83            3               1       1    benign
84            7               1       1    benign
85            7              10       7 malignant
86            4              10      10 malignant
87            4               4       1 malignant
88            6               8       3 malignant
89            3               1       1    benign
90            2               1       1    benign
91            3               1       1    benign
92            1               1       1    benign
93            3               1       1    benign
94            2               1       1    benign
95            3               1       1    benign
96            3               1       1    benign
97            1               1       1    benign
98            3               1       1    benign
99            2               9      10 malignant
100           7               9       4 malignant
101           3              10       2 malignant
102           2               5       1 malignant
103           3               1       1    benign
104           7               1       1 malignant
105           8               8       8 malignant
106           3               2       7 malignant
107           4               1       1 malignant
108           5               7       1 malignant
109           2               3       1    benign
110           7               8       3 malignant
111           5               3       2    benign
112           3               1       1 malignant
113           7               3       3 malignant
114           8               1       1 malignant
115           3               1       1    benign
116           1               1       1    benign
117           3               2       1    benign
118           7               5       8 malignant
119           1               1       1    benign
120           3               1       1    benign
121           3               1       1    benign
122           3               1       1    benign
123           5               3       3 malignant
```

```
124           5               3       1 malignant
125           8              10       1 malignant
126           2               1       1    benign
127           7               5       5 malignant
128           3               1       1    benign
129           1               6       2 malignant
130           1               1       1    benign
131           2               1       1    benign
132           3               1       1    benign
133           3               6       3 malignant
134           2               2       1    benign
135           2               1       1    benign
136           3               3       1    benign
137           2               1       1    benign
138           1               1       1    benign
139           2               1       1    benign
140           2               1       1    benign
141           1               1       1    benign
142           1               1       1    benign
143           4               3       3 malignant
144           1               1       1    benign
145           2               1       1    benign
146           2               1       1    benign
147           4               1       1 malignant
148           2               1       1    benign
149           5               8       1    benign
150           7               8       7 malignant
151           3               1       1    benign
152           5               4       3 malignant
153           8              10       1 malignant
154           1               1       1    benign
155           1               1       1    benign
156           3               1       1 malignant
157           2               1       1    benign
158           3               1       1    benign
159           1               1       1    benign
160           7              10       6 malignant
161           5               7       2 malignant
162           3               2       1    benign
163           3               1       1    benign
164           1               1       7    benign
165           3               1       1    benign
166           3               2       1    benign
```

```
167           3              10       3 malignant
168           3               1      10 malignant
169           3               1       1    benign
170           1               1       1    benign
171           1               1       1    benign
172           3               1       1    benign
173           2               1       1    benign
174          10              10       7 malignant
175           6               1       1 malignant
176           5               7       1 malignant
177           3               1       1    benign
178           5              10       3 malignant
179           3               1       1    benign
180           3               1       1 malignant
181           3               1       1    benign
182           1               1       1    benign
183           3               1       1    benign
184           7               8       1 malignant
185           5               1       1 malignant
186           3               1       1    benign
187           7              10       1 malignant
188           7               7      10 malignant
189           9              10       1 malignant
190           3               1       1    benign
191           7              10       1 malignant
192           4              10       3 malignant
193           2               1       1    benign
194           3               1       1    benign
195           3               1       1    benign
196           3               1       1    benign
197           7               8       2    benign
198           3               1       1    benign
199           1               1       1    benign
200           2               1       1    benign
201           7               8       3 malignant
202           8               1       1 malignant
203           3               1       1    benign
204           3               1       1    benign
205           3               1       1    benign
206           7              10       5 malignant
207           3               5       1 malignant
208           3               1       1    benign
209           3               1       1    benign
```

```
210          3               1       1    benign
211          8              10       6 malignant
212          7               7       1 malignant
213          3               1       1    benign
214          7              10       4 malignant
215         10               6       1 malignant
216          5              10       2 malignant
217          2               1       1    benign
218          3               1       1    benign
219          8              10       2 malignant
220          3               1       1    benign
221          3               1       1    benign
222          9              10       1 malignant
223          2               1       1 malignant
224          7               4       1 malignant
225          7               9       2 malignant
226          2               1       1    benign
227          8               9       1 malignant
228          7               7       1 malignant
229          3               1       1    benign
230          9              10       1 malignant
231          7               6       1 malignant
232          8               9       2 malignant
233          4               3       1    benign
234          4               1       1 malignant
235          3               6       1    benign
236          3               1       1    benign
237          4               8      10 malignant
238          4              10       4 malignant
239          3              10      10 malignant
240          5               3       2 malignant
241          2               3       1    benign
242          3               1       1    benign
243          3               1       1    benign
244          5               1       1    benign
245          3               1       1    benign
246          3               1       1    benign
247          7               8       1 malignant
248          3               3       1 malignant
249          3               6       1    benign
250          3               1       1    benign
251          1               1       1    benign
252          5               3       3 malignant
```

```
253      3          5    3    benign
254      7          3    3 malignant
255      3          3    1 malignant
256      3          6    1 malignant
257      1          1    1    benign
258      2          1    1    benign
259      3          1    1    benign
260      3          4    1    benign
261      5          1    3 malignant
262     10          6    5 malignant
263      7          8    1 malignant
264      5          5    1 malignant
265      5          3    3 malignant
266      3          2    1    benign
267      4          3    2 malignant
268      7          1    1 malignant
269      8          7    8 malignant
270      3          1    1    benign
271      3          9    2 malignant
272      3          1    1    benign
273      7          1    1 malignant
274      3          3    1 malignant
275      3          2    1    benign
276      2          1    1    benign
277      2          1    1    benign
278      2          1    1    benign
279      3          1    1    benign
280      3          3    8 malignant
281      3          1    1    benign
282      3          1    1    benign
283      5          6    1 malignant
284      5          3    1 malignant
285      3          8    2 malignant
286     10          7    3 malignant
287      4         10   10 malignant
288      2          1    1    benign
289      5         10    1 malignant
290      4         10    4 malignant
291      1          1    1    benign
292      3          1    1    benign
293      6         10    1 malignant
294      2          3    1 malignant
295      2          1    1    benign
```

```
296           7               4       1 malignant
297           4               7       1    benign
298           2               3       1    benign
299           1               1       1    benign
300           7               7       2 malignant
301           7              10       1 malignant
302           3               1       1    benign
303           7              10      10 malignant
304           3               1       1    benign
305           3               3       1 malignant
306           3              10       4 malignant
307           3               1       1    benign
308           3               1       1    benign
309           8               8       4 malignant
310           5               1       1    benign
311           2               1       1    benign
312           1               1       1    benign
313           3               5       1 malignant
314           1               1       1    benign
315           2               1       1    benign
316           4               9       1    benign
317           4               3       1 malignant
318           8               9       1 malignant
319           3               1       1    benign
320           7               3       1    benign
321           7               4       6 malignant
322           3               1       1    benign
323           3               1       1    benign
324           4               1       1 malignant
325           3               1       1    benign
326           2               3       1    benign
327           5               4       1 malignant
328           2               1       1    benign
329           3              10       1 malignant
330           7               1       1 malignant
331           6               1       1 malignant
332           3               1       2    benign
333           2               2       1    benign
334           4               3       1 malignant
335           3               4       2 malignant
336           1               1       1    benign
337           3               4       1 malignant
338           3               1       1    benign
```

```
339           2               1       1    benign
340           4               3       1 malignant
341           7               6       1 malignant
342           3               1       1    benign
343           1               1       1    benign
344           1               1       1    benign
345           9               5       3 malignant
346           1               1       1    benign
347           1               3       1    benign
348           1               3       1    benign
349           3               3       1 malignant
350           7               6       1 malignant
351           1               1       1    benign
352           3               1       1    benign
353           4               6       1    benign
354           4               9       4 malignant
355           2               1       1    benign
356           2               2       1    benign
357           3               3       3 malignant
358           7               3       8 malignant
359           4              10       3 malignant
360           3               5       3 malignant
361           8              10      10 malignant
362           5               1       4 malignant
363           2               1       1    benign
364           2               1       1    benign
365           3               1       1    benign
366           2               1       1    benign
367           7              10       7 malignant
368           8              10       3 malignant
369           1               1       1    benign
370           2               1       1    benign
371           2               1       1    benign
372           1               1       1    benign
373           2               1       1    benign
374           2               1       1    benign
375           2               1       1    benign
376           1               1       1    benign
377           2               1       1    benign
378           2               1       1    benign
379           2               2       1    benign
380           3               1       1    benign
381           1               1       1    benign
```

```
382           7               8       4 malignant
383           3               2       1    benign
384           1               1       1    benign
385           1               1       1    benign
386           1               2       3    benign
387           7               1       1 malignant
388           3               1       1    benign
389           2               2       1    benign
390           2               2       1    benign
391           2               1       1    benign
392           7               9       1 malignant
393           2               1       1    benign
394           1               1       1    benign
395           2               1       1    benign
396           2               1       1    benign
397           3               1       1    benign
398           1               1       1    benign
399           2               2       1    benign
400           1               1       1    benign
401           9               3       8 malignant
402           1               1       1    benign
403           2               1       1    benign
404           1               1       1    benign
405           1               2       1    benign
406           2               1       1    benign
407           2               1       1    benign
408           2               1       1    benign
409           3               1       1    benign
410           2               1       1    benign
411           2               1       1    benign
412           2               1       1    benign
413           8               5       1 malignant
414           3               1       1    benign
415           6               6       1 malignant
416           3               5       1    benign
417           7               2       1 malignant
418           2               1       1    benign
419           3               2       2    benign
420           1               1       1    benign
421           3               1       1    benign
422           8               2       1 malignant
423           3               3       1    benign
424           2               1       1    benign
```

```
425           1               1       1    benign
426          10              10       1 malignant
427           1               1       1    benign
428           5              10       1 malignant
429           2               1       1    benign
430           2               1       1    benign
431           2               2       1    benign
432           3               2       1    benign
433           2               2       1    benign
434           1               1       1    benign
435           4               2       1    benign
436           5               1       1 malignant
437           2               8       1 malignant
438           1               1       1    benign
439           1               1       1    benign
440           1               1       1    benign
441          10               1       1 malignant
442           1               1       1    benign
443           1               1       1    benign
444           1               1       1    benign
445           2               1       1    benign
446           1               1       1    benign
447           1               1       1    benign
448           1               1       1    benign
449           1               1       1    benign
450           8              10       1 malignant
451           2               1       1    benign
452           1               1       1    benign
453           1               1       1    benign
454          10               7       1 malignant
455           1               1       1    benign
456           1               1       2 malignant
457           8               6       1 malignant
458          10              10       1 malignant
459           1               1       1    benign
460           1               1       1    benign
461           1               1       1    benign
462           1               1       1    benign
463           1               1       1    benign
464           1               2       1    benign
465           1               1       1    benign
466           7              10       3 malignant
467           9               7       1 malignant
```

```
468           7               6       2 malignant
469           1               1       1    benign
470           2               1       1    benign
471           2               1       1    benign
472           1               1       1    benign
473           1               1       1    benign
474           1               1       1    benign
475           1               1       1    benign
476           1               1       1    benign
477           1               1       1    benign
478           1               1       1    benign
479           1               1       1    benign
480           7               5       1 malignant
481           1               1       1    benign
482           1               1       1    benign
483          10              10      10 malignant
484           9              10       1 malignant
485           1               1       1    benign
486           1               1       1    benign
487           2               1       1    benign
488           8               1       5 malignant
489           3               4       1 malignant
490           4               1       1 malignant
491           1               1       1    benign
492           7               1       1 malignant
493           2               1       1    benign
494           6               5       2 malignant
495           2               1       1    benign
496           2               1       1    benign
497           1               1       1    benign
498           1               1       1    benign
499           2               1       1    benign
500           2               1       1    benign
501           3               1       1    benign
502           2               1       1    benign
503           2               1       1    benign
504           3               1       1    benign
505           1               1       1    benign
506           1               1       1    benign
507           4               8       7 malignant
508           1               1       1    benign
509           1               1       1    benign
510           1               1       1    benign
```

```
511           1               1       1    benign
512           2               1       1    benign
513           1               1       1    benign
514           2               1       1    benign
515           8              10       2 malignant
516           9              10       1 malignant
517           1               1       1    benign
518           2               1       1    benign
519           1               1       1    benign
520           9               1       1 malignant
521           1               1       1    benign
522           1               1       1    benign
523           7               3       1 malignant
524           5               3       1 malignant
525           2               1       1    benign
526           1               1       1    benign
527           1               1       1    benign
528           3               1       1    benign
529           1               1       1    benign
530           2               1       1    benign
531           6               9       1 malignant
532           2               1       1    benign
533           3               1       1    benign
534           2               1       1    benign
535           2               1       1    benign
536           3               1       1    benign
537           3               1       1    benign
538           3               1       1    benign
539           2               1       1    benign
540           2               1       1    benign
541           2               1       1    benign
542           1               1       1    benign
543           1               1       1    benign
544           2               1       1    benign
545           2               1       1    benign
546           2               1       1    benign
547           7              10       1 malignant
548           1               1       1    benign
549           1               1       1    benign
550           7               8       2 malignant
551           2               1       1    benign
552           3               1       1    benign
553           4               2       1    benign
```

```
554           2               1       2    benign
555           1               1       1    benign
556           4               8       1    benign
557           2               1       1    benign
558           1               1       1    benign
559           2               1       1    benign
560           2               1       1    benign
561           3               1       1    benign
562           3               1       1    benign
563           3               1       1    benign
564           2               1       1    benign
565           3               2       1    benign
566          10              10       1 malignant
567           3               1       1    benign
568           2               1       1    benign
569           2               5       2 malignant
570          10               3       1 malignant
571           8               2       1 malignant
572           9              10       2 malignant
573           2               1       1    benign
574           2               1       1    benign
575           7               7       1 malignant
576           3               1       1    benign
577           2               1       1    benign
578           2               1       1    benign
579           2               1       1    benign
580           3               1       1    benign
581           2               1       1    benign
582           7               5       1 malignant
583           6              10       1 malignant
584           1               1       1    benign
585           1               1       1    benign
586           1               1       1    benign
587          10              10       1 malignant
588           2               2       1    benign
589           4               1       1 malignant
590           1               1       1    benign
591          10               1       1 malignant
592           7               6       1 malignant
593           4               1       1 malignant
594           1               1       1    benign
595           7               1       1 malignant
596           2               1       1    benign
```

```
597           2               1       1    benign
598           3               1       1    benign
599           2               1       1    benign
600           1               1       1    benign
601           2               1       1    benign
602           2               1       1    benign
603           2               1       1    benign
604           8              10       1 malignant
605           8               1       2 malignant
606           7               8       3 malignant
607           1               1       1    benign
608           1               1       1    benign
609          10               1       1 malignant
610           1               1       1    benign
611           7               1       2 malignant
612           8               5       1 malignant
613          10              10      10 malignant
614           2               1       1    benign
615           2               1       1    benign
616           2               1       1    benign
617           2               1       1    benign
618           1               1       1    benign
619           2               1       1    benign
620           2               1       1    benign
621           2               1       1    benign
622           6               1       1    benign
623           2               1       1    benign
624           1               1       1    benign
625           2               1       1    benign
626           1               1       1    benign
627           7               7       3 malignant
628           1               1       1    benign
629           1               1       1    benign
630           1               1       1    benign
631           1               1       1    benign
632           2               1       1    benign
633           1               1       1    benign
634           5              10       1 malignant
635           1               1       1    benign
636           1               1       1    benign
637          10              10       3 malignant
638           2               1       1    benign
639           1               1       1    benign
```

```
640           1               1       1    benign
641           1               1       1    benign
642           2               1       1    benign
643           2               1       1    benign
644           1               1       1    benign
645           1               1       1    benign
646           2               1       1    benign
647           1               1       1    benign
648           1               1       1    benign
649          10              10      10 malignant
650           2               1       1    benign
651           1               1       1    benign
652           2               1       1    benign
653           2               2       1    benign
654           2               1       1    benign
655           3               1       1    benign
656           2               1       1    benign
657           2               1       1    benign
658           3               6       1    benign
659           7               2       3 malignant
660           1               1       1    benign
661           2               1       1    benign
662           3               1       1    benign
663           2               1       1    benign
664           2               1       1    benign
665           2               1       1    benign
666           1               1       1    benign
667           1               1       2    benign
668           3               1       1    benign
669           7              10       3 malignant
670           7              10       1 malignant
671           7               4       1 malignant
672           3               1       1    benign
673           3               1       1    benign
674           1               1       1    benign
675           2               1       1    benign
676           1               1       1    benign
677           2               1       1    benign
678           1               1       1    benign
679           1               1       1    benign
680           1               1       1    benign
681          10              10       7 malignant
682           5               6       3 malignant
```

```
683            3               2       1    benign
684            1               1       1    benign
685            1               1       1    benign
686            1               1       1    benign
687            1               1       1    benign
688            2               3       1    benign
689            1               1       1    benign
690            1               1       8    benign
691            1               1       1    benign
692            4               4       1 malignant
693            1               1       1    benign
694            2               1       2    benign
695            1               1       1    benign
696            1               1       1    benign
697            8              10       2 malignant
698           10               6       1 malignant
699           10               4       1 malignant
```

```r
sum(is.na(df))
```

```
[1] 16
```

```r
colSums(is.na(df))
```

```
   Cl.thickness       Cell.size      Cell.shape   Marg.adhesion    Epith.c.size 
              0               0               0               0               0 
    Bare.nuclei     Bl.cromatin Normal.nucleoli         Mitoses           Class 
             16               0               0               0               0 
```

```r
df <- na.omit(df)
df$Cl.thickness <- as.numeric(df$Cl.thickness)
df$Cell.size <- as.numeric(df$Cell.size)
df$Cell.shape <- as.numeric(df$Cell.shape)
df$Marg.adhesion <- as.numeric(df$Marg.adhesion)
df$Epith.c.size <- as.numeric(df$Epith.c.size)
df$Bare.nuclei <- as.numeric(df$Bare.nuclei)
df$Bl.cromatin <- as.numeric(df$Bl.cromatin)
df$Normal.nucleoli <- as.numeric(df$Normal.nucleoli)
df$Mitoses <- as.numeric(df$Mitoses)

set.seed(1234)
```

```
index <- sample(nrow(df), 0.7*nrow(df))
train <- df[index,]
test <- df[-index,]
```

Then, we use the nine variables to perform logistic regression for classification, where the class is the dependent variable. We use the summary function to examine the details of the fit result. Next, we use the test data for prediction and set type = 'response' to see the probability of the malignant class.

Next, we set a threshold for the probability: if the probability is greater than 0.5, we label it as 'malignant', and if it is less than 0.5, we label it as 'benign'. Then, we convert these 'benign' and 'malignant' labels into factors for further classification, as we will use the table function.

```
fit <- glm(Class ~ ., data = train, family = binomial())
summary(fit)
```

```
Call:
glm(formula = Class ~ ., family = binomial(), data = train)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)    
(Intercept)     -9.69151    1.29558  -7.480 7.41e-14 ***
Cl.thickness     0.48040    0.15234   3.154  0.00161 ** 
Cell.size        0.05705    0.29262   0.195  0.84541    
Cell.shape       0.13100    0.31630   0.414  0.67874    
Marg.adhesion    0.40724    0.14038   2.901  0.00372 ** 
Epith.c.size    -0.03252    0.18088  -0.180  0.85731    
Bare.nuclei      0.44746    0.11178   4.003 6.26e-05 ***
Bl.cromatin      0.48273    0.19220   2.512  0.01202 *  
Normal.nucleoli  0.23560    0.12904   1.826  0.06788 .  
Mitoses          0.66368    0.28592   2.321  0.02028 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 629.040  on 477  degrees of freedom
Residual deviance:  80.061  on 468  degrees of freedom
AIC: 100.06
```

```
Number of Fisher Scoring iterations: 8
```

```r
prob <- predict(fit,test,type="response")

prediction <- factor(prob>0.5, levels=c(FALSE,TRUE), labels = c("benign","malignant"))
prediction
```

```
        3         5         6         8         9        16        17        21
   benign    benign malignant    benign    benign    benign    benign malignant
       25        27        28        30        33        38        39        46
   benign    benign    benign    benign malignant    benign malignant    benign
       49        50        57        58        59        64        71        74
   benign malignant malignant malignant malignant    benign    benign malignant
       75        82        83        84        87        92        93        94
malignant    benign    benign    benign malignant    benign    benign    benign
       95        98       111       112       113       114       119       126
   benign    benign    benign malignant malignant malignant    benign    benign
      130       131       132       134       135       136       139       142
   benign    benign    benign    benign    benign    benign    benign    benign
      143       152       155       157       163       173       178       179
malignant malignant    benign    benign    benign    benign malignant    benign
      182       186       191       193       196       199       203       206
   benign    benign malignant    benign    benign    benign    benign malignant
      210       215       228       231       232       239       240       244
   benign malignant malignant malignant malignant malignant malignant    benign
      245       246       253       254       255       261       272       274
   benign    benign malignant malignant malignant malignant    benign    benign
      275       277       284       287       288       291       297       299
   benign    benign malignant malignant    benign    benign malignant    benign
      301       304       311       312       313       314       317       319
malignant    benign    benign    benign malignant    benign malignant    benign
      326       329       332       335       339       341       343       344
   benign malignant    benign malignant    benign malignant    benign    benign
      346       351       354       368       369       373       381       383
   benign    benign malignant malignant    benign    benign    benign    benign
      395       405       407       408       413       416       417       425
   benign    benign    benign    benign malignant    benign malignant    benign
      426       429       430       431       434       437       448       450
malignant    benign    benign    benign    benign malignant    benign malignant
      452       453       463       466       467       469       470       479
   benign    benign    benign malignant malignant    benign    benign    benign
      480       481       488       498       502       505       509       511
```

```
malignant    benign malignant    benign    benign    benign    benign    benign
      519       520       522       524       529       532       536       539
   benign malignant    benign malignant    benign    benign    benign    benign
      540       546       548       551       554       557       560       564
   benign    benign    benign    benign    benign    benign    benign    benign
      566       570       573       574       575       578       579       580
malignant malignant    benign    benign malignant    benign    benign    benign
      583       596       598       600       601       603       605       606
malignant    benign    benign    benign    benign    benign malignant malignant
      608       609       611       613       628       629       632       638
   benign malignant malignant malignant    benign    benign    benign    benign
      640       641       646       648       655       656       658       660
   benign    benign    benign    benign    benign    benign    benign    benign
      665       667       668       670       672       676       678       679
   benign    benign    benign malignant    benign    benign    benign    benign
      680       681       684       686       694
   benign malignant    benign    benign    benign
Levels: benign malignant
```

Finally, we generate a table comparing the actual classification with the predicted classification, which allows us to view the final results. Based on this table, we calculate four key metrics to evaluate the model's performance: Accuracy, Precision, Recall, and F1 Score. From these metrics, we observe that the result is generally acceptable and high.

```
perf1 <- table(test$Class, prediction, dnn=c("Actual","Predicted"))
perf1
```

```
           Predicted
Actual       benign malignant
  benign        140         2
  malignant       3        60
```

```
TN1 <- perf1[1, 1]  # True Negatives
TP1 <- perf1[2, 2]  # True Positives

FP1 <- perf1[1, 2]  # False Positives
FN1 <- perf1[2, 1]  # False Negatives


#  (Accuracy)
accuracy1 <- (TP1 + TN1) / sum(perf1)
```

```
# (Precision)
precision1 <- TP1 / (TP1 + FP1)

#  (Recall)
recall1 <- TP1 / (TP1 + FN1)


spcificity1 <- TN1/(TN1 + FP1)

list(
  Accuracy1 = accuracy1,
  Precision1 = precision1,
  Recall1 = recall1,
  Spcificity1 = spcificity1
)
```

```
$Accuracy1
[1] 0.9756098

$Precision1
[1] 0.9677419

$Recall1
[1] 0.952381

$Spcificity1
[1] 0.9859155
```

Decision Tree

Next, we move on to the Decision Tree method. We set the parameters to use information gain instead of the Gini index. By using the print and summary functions, we can examine the details of the decision tree.

```
library("rpart")
dtree <- rpart(Class ~ .,data= train, method="class",parms=list(split="information"))

print(dtree)
```

```
n= 478
```

```
node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 478 176 benign (0.63179916 0.36820084)
   2) Cell.size< 2.5 279   8 benign (0.97132616 0.02867384) *
   3) Cell.size>=2.5 199  31 malignant (0.15577889 0.84422111)
     6) Cell.size< 4.5 69  28 malignant (0.40579710 0.59420290)
      12) Bare.nuclei< 2.5 25   3 benign (0.88000000 0.12000000) *
      13) Bare.nuclei>=2.5 44   6 malignant (0.13636364 0.86363636) *
     7) Cell.size>=4.5 130   3 malignant (0.02307692 0.97692308) *
```

```
summary(dtree)
```

```
Call:
rpart(formula = Class ~ ., data = train, method = "class", parms = list(split = "information
  n= 478

          CP nsplit rel error    xerror       xstd
1 0.77840909      0 1.0000000 1.0000000 0.05991467
2 0.05397727      1 0.2215909 0.2556818 0.03627635
3 0.01000000      3 0.1136364 0.1250000 0.02602958

Variable importance
      Cell.size      Cell.shape    Epith.c.size     Bare.nuclei     Bl.cromatin 
             22              18              16              15              15 
Normal.nucleoli   Marg.adhesion 
             14               1 

Node number 1: 478 observations,    complexity param=0.7784091
  predicted class=benign     expected loss=0.3682008  P(node) =1
    class counts:   302   176
   probabilities: 0.632 0.368 
  left son=2 (279 obs) right son=3 (199 obs)
  Primary splits:
      Cell.size    < 2.5 to the left,  improve=192.1333, (0 missing)
      Cell.shape   < 2.5 to the left,  improve=188.4330, (0 missing)
      Bare.nuclei  < 2.5 to the left,  improve=168.9610, (0 missing)
      Bl.cromatin  < 3.5 to the left,  improve=163.3498, (0 missing)
      Epith.c.size < 2.5 to the left,  improve=157.6645, (0 missing)
  Surrogate splits:
      Cell.shape      < 2.5 to the left,  agree=0.918, adj=0.804, (0 split)
      Epith.c.size    < 2.5 to the left,  agree=0.897, adj=0.754, (0 split)
      Bl.cromatin     < 3.5 to the left,  agree=0.868, adj=0.683, (0 split)
      Bare.nuclei     < 2.5 to the left,  agree=0.860, adj=0.663, (0 split)
      Normal.nucleoli < 2.5 to the left,  agree=0.860, adj=0.663, (0 split)

Node number 2: 279 observations
  predicted class=benign     expected loss=0.02867384  P(node) =0.583682
    class counts:   271     8
   probabilities: 0.971 0.029

Node number 3: 199 observations,    complexity param=0.05397727
  predicted class=malignant  expected loss=0.1557789  P(node) =0.416318
    class counts:    31   168
   probabilities: 0.156 0.844
  left son=6 (69 obs) right son=7 (130 obs)
  Primary splits:
      Cell.size     < 4.5 to the left,  improve=25.22106, (0 missing)
      Bare.nuclei   < 2.5 to the left,  improve=24.73323, (0 missing)
      Cell.shape    < 2.5 to the left,  improve=21.68253, (0 missing)
      Marg.adhesion < 5.5 to the left,  improve=19.06313, (0 missing)
      Bl.cromatin   < 3.5 to the left,  improve=18.80490, (0 missing)
  Surrogate splits:
      Cell.shape      < 4.5 to the left,  agree=0.794, adj=0.406, (0 split)
      Epith.c.size    < 3.5 to the left,  agree=0.759, adj=0.304, (0 split)
      Marg.adhesion   < 2.5 to the left,  agree=0.749, adj=0.275, (0 split)
      Bl.cromatin     < 3.5 to the left,  agree=0.749, adj=0.275, (0 split)
      Normal.nucleoli < 3.5 to the left,  agree=0.724, adj=0.203, (0 split)

Node number 6: 69 observations,    complexity param=0.05397727
  predicted class=malignant  expected loss=0.4057971  P(node) =0.1443515
    class counts:    28    41
   probabilities: 0.406 0.594
  left son=12 (25 obs) right son=13 (44 obs)
  Primary splits:
      Bare.nuclei     < 2.5 to the left,  improve=19.896530, (0 missing)
      Cell.shape      < 2.5 to the left,  improve= 9.967081, (0 missing)
      Cl.thickness    < 9   to the left,  improve= 8.481164, (0 missing)
      Normal.nucleoli < 2.5 to the left,  improve= 7.230538, (0 missing)
      Marg.adhesion   < 5.5 to the left,  improve= 7.094551, (0 missing)
  Surrogate splits:
      Cell.shape      < 2.5 to the left,  agree=0.812, adj=0.48, (0 split)
      Bl.cromatin     < 1.5 to the left,  agree=0.739, adj=0.28, (0 split)
      Marg.adhesion   < 1.5 to the left,  agree=0.710, adj=0.20, (0 split)
      Cl.thickness    < 2.5 to the left,  agree=0.681, adj=0.12, (0 split)
      Normal.nucleoli < 2.5 to the left,  agree=0.681, adj=0.12, (0 split)

Node number 7: 130 observations
  predicted class=malignant  expected loss=0.02307692  P(node) =0.2719665
    class counts:     3   127
   probabilities: 0.023 0.977

Node number 12: 25 observations
  predicted class=benign     expected loss=0.12  P(node) =0.05230126
    class counts:    22     3
   probabilities: 0.880 0.120

Node number 13: 44 observations
  predicted class=malignant  expected loss=0.1363636  P(node) =0.09205021
    class counts:     6    38
   probabilities: 0.136 0.864
```

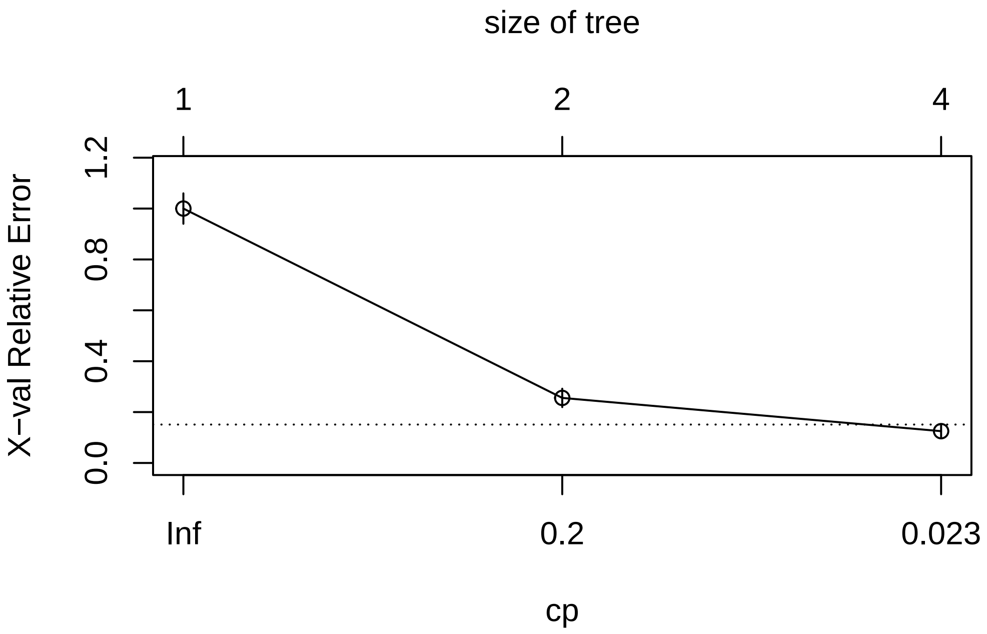
In order to simplify the model, improve calculation efficiency, and enhance stability, we use the prune function to help optimize the decision tree.

By examining the complexity parameter table, we can determine the optimal pruning point for the decision tree, which shows different values of cp, the number of branches in the tree (nsplit), the relative error (rel error), the cross-validation error (xerror), and its standard deviation (xstd).

The general approach is to choose the value of **CP** that minimizes the **cross-validation error (xerror)**.

In that case, we will choose the 0.01 for the CP value. After pruning the decision tree, we can use fancyRpartPlot function to visualize the decision tree and its structure clearly.

```r
plotcp(dtree)
```



```
dtree$cptable
```

```
          CP nsplit rel error    xerror       xstd
1 0.77840909      0 1.0000000 1.0000000 0.05991467
2 0.05397727      1 0.2215909 0.2556818 0.03627635
3 0.01000000      3 0.1136364 0.1250000 0.02602958
```

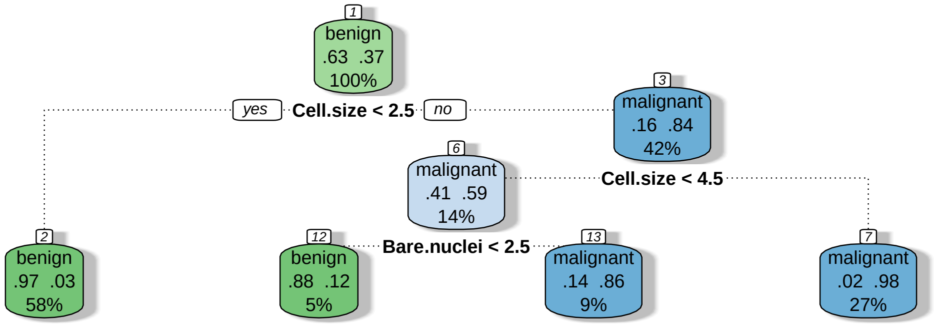
```
dtree.pruned <- prune(dtree,cp=0.01)

library(rattle)
```

```
Loading required package: tibble
```

```
Loading required package: bitops
```

```
Rattle: A free graphical interface for data science with R.
Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.
Type 'rattle()' to shake, rattle, and roll your data.
```

```
fancyRpartPlot(dtree.pruned, sub="Classification Tree")
```



Classification Tree

After pruning the decision tree, we can use the test data to make predictions. Similar to logistic regression, we can use the table() function to compare the actual classifications with the predicted classifications.

```
dtree.pred <- predict(dtree.pruned,test,type="class")
perf2 <- table(test$Class,dtree.pred,dnn=c("Actual","Predicted"))
perf2
```

```
           Predicted
Actual       benign malignant
  benign        138         4
  malignant       6        57
```

```
TN2 <- perf2[1, 1]  # True Negatives
FP2 <- perf2[1, 2]  # False Positives
FN2 <- perf2[2, 1]  # False Negatives
TP2 <- perf2[2, 2]  # True Positives

#  (Accuracy)
accuracy2 <- (TP2 + TN2) / sum(perf2)
```

```
#  (Precision)
precision2 <- TP2 / (TP2 + FP2)

#  (Recall)
recall2 <- TP2 / (TP2 + FN2)

spcificity2 <- TN2/(TN2 + FP2)

list(
  Accuracy2 = accuracy2,
  Precision2 = precision2,
  Recall2 = recall2,
  Spcificity2 =  spcificity2
)
```

```
$Accuracy2
[1] 0.9512195

$Precision2
[1] 0.9344262

$Recall2
[1] 0.9047619

$Spcificity2
[1] 0.971831
```

RandomForestModel

RandomForest

Now, let's move on to the random forest method. We use the randomForest function to build the model and the importance function to evaluate and rank the variables based on their importance.

```
library(randomForest)
```

```
randomForest 4.7-1.2
```

```
Type rfNews() to see new features/changes/bug fixes.
```

```
Attaching package: 'randomForest'


The following object is masked from 'package:rattle':

    importance
```

```
set.seed(1234)
fit.forest <- randomForest(Class ~ ., data=train,importance= TRUE)
fit.forest
```

```

Call:
 randomForest(formula = Class ~ ., data = train, importance = TRUE)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 2.93%
Confusion matrix:
          benign malignant class.error
benign       293         9  0.02980132
malignant      5       171  0.02840909
```

```
var_importance <- fit.forest$importance
print(var_importance)
```

```
                     benign   malignant MeanDecreaseAccuracy MeanDecreaseGini
Cl.thickness    0.042688778 0.026907087          0.036757666         9.794852
Cell.size       0.090035717 0.087705808          0.088954598        58.635963
Cell.shape      0.021820598 0.073103200          0.040661838        49.754466
Marg.adhesion   0.011159602 0.032511581          0.018912111         8.373530
Epith.c.size    0.019658049 0.005981130          0.014562683        16.814313
Bare.nuclei     0.074536451 0.069801310          0.072818091        36.621347
Bl.cromatin     0.018291504 0.038034280          0.025639056        25.179804
Normal.nucleoli 0.022615855 0.017624908          0.020754772        14.177153
Mitoses         0.005865146 0.003703546          0.004991195         2.015803
```

Use the test data

```r
forest.pred <- predict(fit.forest,test)
perf3 <- table(test$Class,forest.pred,dnn=c("Actual","Predicted"))
perf3
```

```
           Predicted
Actual      benign malignant
  benign       140         2
  malignant      3        60
```

```r
TN3 <- perf3[1, 1]  # True Negatives
FP3 <- perf3[1, 2]  # False Positives
FN3 <- perf3[2, 1]  # False Negatives
TP3 <- perf3[2, 2]  # True Positives


accuracy3 <- (TP3 + TN3) / sum(perf3)

precision3 <- TP3 / (TP3 + FP3)


recall3 <- TP3 / (TP3 + FN3)
spcificity3 <- TN3/(TN3 + FP3)

list(
  Accuracy3 = accuracy3,
  Precision3 = precision3,
  Recall3 = recall3,
  Spcificity3 = spcificity3
)
```

```
$Accuracy3
[1] 0.9756098

$Precision3
[1] 0.9677419

$Recall3
[1] 0.952381

$Spcificity3
[1] 0.9859155
```

Black-box characteristic

These classification models are really important in real-world applications, as they can have a significant impact on people's decisions. For example, if someone applies for a bank loan and gets rejected, we would want to know the reason. If the decision is based on a logistic regression model or a decision tree model, we can easily understand the exact reason from the coefficients of the logistic regression model or the decision tree plot. However, if the decision is based on a random forest model, how can we get the answer? In random forest, multiple decision trees are used, and the paths of each tree are determined by various factors. As a result, we don't have a clear understanding of the decision path. To address this, there are methods we can use to interpret black box characteristics. One way is by applying explainable artificial intelligence (XAI) techniques to make these decisions more transparent and understandable.

For example, if we randomly generate data for Amy and use the model to calculate her probability of being malignant, and the result shows that her probability of being malignant is nearly 67%, which is greater than 50%, we want to understand how this result is influenced by the nine variables. Specifically, we want to know the contribution of each variable to the final decision. To do this, we use the explainer and plot to clearly visualize how each variable contributes to the model's output.

From the two charts below, we can observe that the results are consistent. The most important factor is the CL thickness, as its contribution is the highest among all the variables

```
library("DALEX")
```

```
Welcome to DALEX (version: 2.4.3).
Find examples and detailed introduction at: http://ema.drwhy.ai/
```

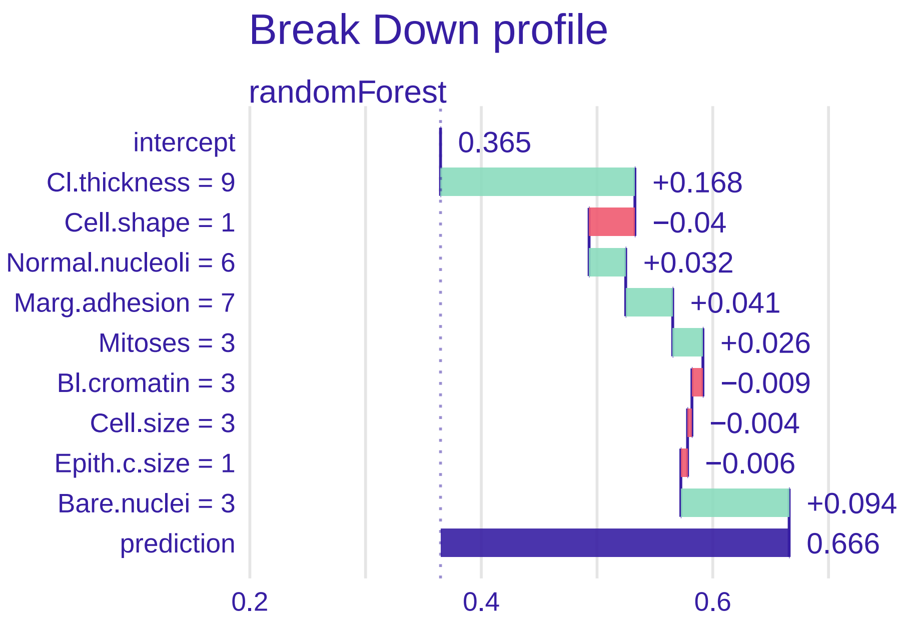
```
Amy <- data.frame(
  Cl.thickness = 9,
  Cell.size = 3,
  Cell.shape = 1,
  Marg.adhesion =7,
  Epith.c.size = 1,
  Bare.nuclei = 3,
  Bl.cromatin = 3,
  Normal.nucleoli = 6,
  Mitoses = 3
)

predict(fit.forest,Amy,type="prob")
```

```
  benign malignant
1  0.334     0.666
attr(,"class")
[1] "matrix" "array"  "votes"
```

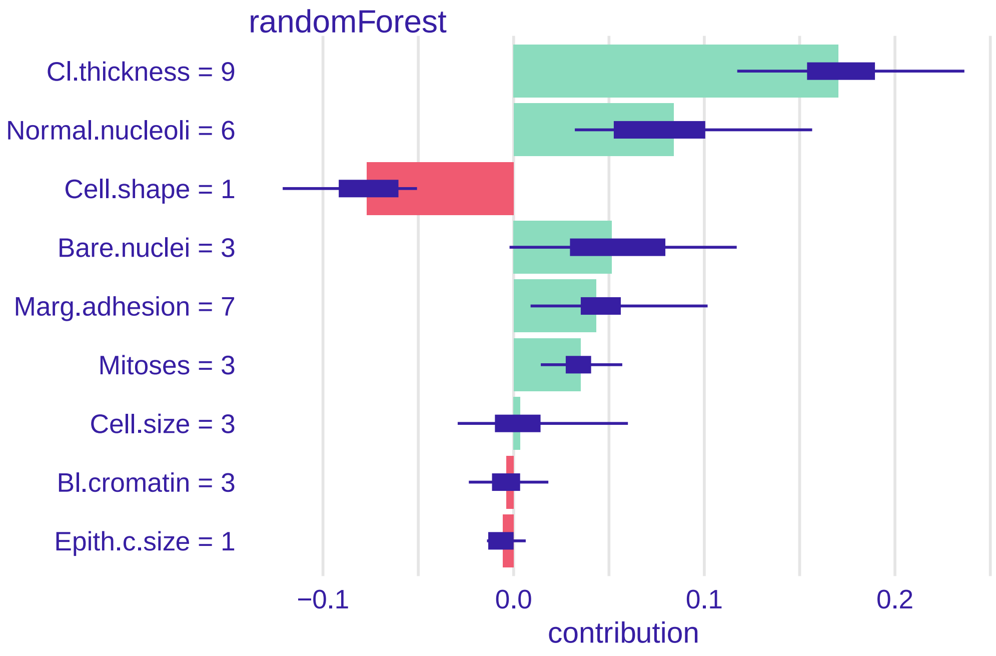
```
explainer_fr_malignant <- explain(fit.forest,data=train,y=train$Class == "malignant",
                      predcit_function = function(m,x) predict(m,x,type = "prob")[,2])
```

```
Preparation of a new explainer is initiated
  -> model label       :  randomForest  (  default  )
  -> data              :  478  rows  10  cols
  -> target variable   :  478  values
  -> predict function  :  yhat.randomForest  will be used (  default  )
  -> predicted values  :  No value for predict function target column. (  default  )
  -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  defaul
  -> model_info        :  Model info detected classification task but 'y' is a logical . Conv
  -> predicted values  :  numerical, min =  0 , mean =  0.3647782 , max =  1
  -> residual function :  difference between y and yhat (  default  )
  -> residuals         :  numerical, min =  -0.364 , mean =  0.003422594 , max =  0.416
  A new explainer has been created!
```

```
rf_pparts1  <- predict_parts(explainer_fr_malignant,new_observation = Amy,type="break_down")

plot(rf_pparts1 )
```



```
rf_pparts2  <- predict_parts(explainer_fr_malignant,new_observation = Amy,type="shap")

plot(rf_pparts2 )
```



Support vector machines

SVM

The results from the Support Vector Machine (SVM) model demonstrate strong performance, as evidenced by the four performance metrics.

```
library(e1071)
set.seed(1234)
fit.svm <- svm(Class ~ ., data=train)
fit.svm
```

```


Call:
svm(formula = Class ~ ., data = train)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  84
```

```
svm.pred <- predict(fit.svm,test)
svm.perf <- table(test$Class,svm.pred,dnn=c("Actual","Predicted"))
svm.perf
```

```
           Predicted
Actual       benign malignant
  benign        138         4
  malignant       1        62
```

```
TN <- svm.perf[1, 1]  # True Negatives
FP <- svm.perf[1, 2]  # False Positives
FN <- svm.perf[2, 1]  # False Negatives
TP <- svm.perf[2, 2]  # True Positives

accuracy <- (TP + TN) / sum(svm.perf)
precision <- TP / (TP + FP)
recall <- TP / (TP + FN)
```

```
spcificity <- TN/(TN + FP)


list(
  Accuracy = accuracy,
  Precision = precision,
  Recall = recall,
  Spcificity =spcificity
)
```

```
$Accuracy
[1] 0.9756098

$Precision
[1] 0.9393939

$Recall
[1] 0.984127

$Spcificity
[1] 0.971831
```

Parameters

There are two key hyperparameters in Support Vector Machines (SVM): gamma and cost. By adjusting these parameters, we can enhance the model's performance. From the code results below, the optimal parameter values are found to be gamma = 0.01 and cost = 1. After updating the default parameters with these values and re-running the model, we observe that the results improve marginally, with one additional benign instance being accurately predicted. While the overall performance remains largely unchanged, it's important to note that tuning parameters in SVM typically leads to improvements in model performance.

```
#SVM
set.seed(1234)
tuned <- tune.svm(Class~.,data=train,gamma=10^(-6:1),cost=10^(-10:10))
tuned
```

```


Parameter tuning of 'svm':

- sampling method: 10-fold cross validation
- best parameters:
 gamma cost
  0.01    1

- best performance: 0.03355496
```

```r
fit.svm1 <- svm(Class~., data=train,gamma=0.01,cost=1)
svm.pred1 <- predict(fit.svm1,na.omit(test))
svm.perf1 <- table(na.omit(test)$Class,svm.pred1,dnn=c("Actual","Predicted"))
svm.perf1
```

```
           Predicted
Actual      benign malignant
  benign       139         3
  malignant      1        62
```

Summary

Among the four models—Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine—the results from all models are generally acceptable, with good performance across the board. However, in practice, the decision to choose the best model depends on the specific context. In this case, while the Support Vector Machine (SVM) appears to offer slightly better performance, the differences among the models are minimal.

From a cancer diagnosis perspective, recall is a critical metric, as it reflects the model's ability to accurately identify malignant cases. Therefore, this metric should be prioritized when selecting the most suitable model. Among the four models, the Support Vector Machine (SVM) demonstrates the highest recall rate, making it the preferred choice for this task.

Logistic regression is a simple and interpretable model that provides probabilities, making it easy to understand and useful for decision-making. It works well with linear data but struggles with nonlinear relationships, limiting its performance in more complex scenarios. Decision trees, on the other hand, can handle nonlinear data and provide a clear, interpretable structure through a tree diagram. They are capable of capturing more complex patterns than logistic regression. Random forests, an ensemble of multiple decision trees, improve upon decision trees by reducing overfitting and improving accuracy. They excel at handling missing data and complex nonlinear relationships, making them more robust and accurate than individual decision trees, but at the cost of interpretability.The Support Vector Machine (SVM) is a widely used and popular method, known for its versatility across various applications. It is particularly useful in scenarios where the number of variables exceeds the number of observations, a common challenge in the pharmaceutical industry. Like Random Forest, SVM's classification principle can be difficult to interpret, as it functions as a black-box model. Additionally, when dealing with large datasets, SVM may not perform as well as Random Forest. However, once a successful model is developed, it is highly effective for classifying new observations.

```
list(
  Accuracy1 = accuracy1,
  Precision1 = precision1,
  Recall1 = recall1,
  spcificity1 = spcificity1
)
```

```
$Accuracy1
[1] 0.9756098

$Precision1
[1] 0.9677419

$Recall1
[1] 0.952381

$spcificity1
[1] 0.9859155
```

```
list(
  Accuracy2 = accuracy2,
  Precision2 = precision2,
  Recall2 = recall2,
  Spcificity2 = spcificity2
)
```

```
$Accuracy2
[1] 0.9512195

$Precision2
[1] 0.9344262

$Recall2
[1] 0.9047619

$Spcificity2
[1] 0.971831
```

```
list(
  Accuracy3 = accuracy3,
  Precision3 = precision3,
  Recall3 = recall3,
  Spcificity3 = spcificity3
)
```

```
$Accuracy3
[1] 0.9756098

$Precision3
[1] 0.9677419

$Recall3
[1] 0.952381

$Spcificity3
[1] 0.9859155
```

```
list(
  Accuracy4 = accuracy,
  Precision4 = precision,
  Recall4 = recall,
  Spcificity4 = spcificity
)
```

```
$Accuracy4
[1] 0.9756098

$Precision4
[1] 0.9393939

$Recall4
[1] 0.984127

$Spcificity4
[1] 0.971831
```