Breast_cancer_calssfication_XGBoost

Xinzhu Li

Introduction

For the dataset Breastcancer classification problem, previously we used 4 methods to solve this problem: Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines(SVM).

Now we use another way from ensemble learning : XGBoost.

XGBoost (Extreme Gradient Boosting) is an advanced **ensemble learning** algorithm that builds multiple decision trees sequentially and improves performance using gradient boosting. It is designed for speed and accuracy, making it a strong choice for structured data like the **Breast Cancer dataset**.

Data preparation

First,we load the data and then we drop the observations that contains missing values. Next, we generate the train dataset and test dataset by using the sample function. After doing that, we set the variables into proper data type.

```
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"
```

```
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

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

```
[1] 16
```

```
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
```

```
df <- na.omit(df)
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
              0               0               0               0               0
```

```
set.seed(1234)
index <- sample(nrow(df), 0.7*nrow(df))
train <- df[index,]
test <- df[-index,]

class(train)
```

```
[1] "data.frame"
```

```
str(train)
```

```
'data.frame':   478 obs. of  10 variables:
 $ Cl.thickness   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 10 4 4 1 8 7 10 3 8 2 ...
 $ Cell.size      : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 4 1 1 1 5 5 10 1 5 2 ...
 $ Cell.shape     : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 4 2 1 1 6 6 10 1 5 2 ...
 $ Marg.adhesion  : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 6 1 1 1 2 10 10 1 5 1 ...
 $ Epith.c.size   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 2 2 2 2 3 5 10 2 2 1 ...
 $ Bare.nuclei    : Factor w/ 10 levels "1","2","3","4",..: 10 1 1 1 10 10 1 1 10 1 ...
 $ Bl.cromatin    : Factor w/ 10 levels "1","2","3","4",..: 2 3 1 2 6 7 8 2 4 7 ...
 $ Normal.nucleoli: Factor w/ 10 levels "1","2","3","4",..: 3 1 1 1 6 9 8 1 3 1 ...
```

```
 $ Mitoses        : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 4 8 1 1 1 ...
 $ Class          : Factor w/ 2 levels "benign","malignant": 2 1 1 1 2 2 2 1 2 1 ...
 - attr(*, "na.action")= 'omit' Named int [1:16] 24 41 140 146 159 165 236 250 276 293 ...
  ..- attr(*, "names")= chr [1:16] "24" "41" "140" "146" ...
```

```
str(train$Class)
```

```
 Factor w/ 2 levels "benign","malignant": 2 1 1 1 2 2 2 1 2 1 ...
```

```
levels(train$Class)
```

```
[1] "benign"    "malignant"
```

```
# [1] "benign"    "malignant"

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

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


str(test)
```

```
'data.frame':   205 obs. of  10 variables:
 $ Cl.thickness   : num  3 4 8 2 2 7 4 7 1 3 ...
```

```
 $ Cell.size      : num  1 1 10 1 1 4 1 3 1 2 ...
 $ Cell.shape     : num  1 1 10 2 1 6 1 2 1 1 ...
 $ Marg.adhesion  : num  1 3 8 1 1 4 1 10 1 1 ...
 $ Epith.c.size   : num  2 2 7 2 2 6 2 5 2 1 ...
 $ Bare.nuclei    : num  2 1 10 1 1 1 1 10 1 1 ...
 $ Bl.cromatin    : num  3 3 9 3 1 4 2 5 3 2 ...
 $ Normal.nucleoli: num  1 1 7 1 1 3 1 4 1 1 ...
 $ Mitoses        : num  1 1 1 1 5 1 1 4 1 1 ...
 $ Class          : Factor w/ 2 levels "benign","malignant": 1 1 2 1 1 2 1 2 1 1 ...
 - attr(*, "na.action")= 'omit' Named int [1:16] 24 41 140 146 159 165 236 250 276 293 ...
  ..- attr(*, "names")= chr [1:16] "24" "41" "140" "146" ...
```

```
str(train)
```

```
'data.frame':   478 obs. of  10 variables:
 $ Cl.thickness   : num  10 4 4 1 8 7 10 3 8 2 ...
 $ Cell.size      : num  4 1 1 1 5 5 10 1 5 2 ...
 $ Cell.shape     : num  4 2 1 1 6 6 10 1 5 2 ...
 $ Marg.adhesion  : num  6 1 1 1 2 10 10 1 5 1 ...
 $ Epith.c.size   : num  2 2 2 2 3 5 10 2 2 1 ...
 $ Bare.nuclei    : num  10 1 1 1 10 10 1 1 10 1 ...
 $ Bl.cromatin    : num  2 3 1 2 6 7 8 2 4 7 ...
 $ Normal.nucleoli: num  3 1 1 1 6 9 8 1 3 1 ...
 $ Mitoses        : num  1 1 1 1 1 4 8 1 1 1 ...
 $ Class          : Factor w/ 2 levels "benign","malignant": 2 1 1 1 2 2 2 1 2 1 ...
 - attr(*, "na.action")= 'omit' Named int [1:16] 24 41 140 146 159 165 236 250 276 293 ...
  ..- attr(*, "names")= chr [1:16] "24" "41" "140" "146" ...
```

XGBoost model

It first extracts feature variables as well as the label variables. And then converts the target labels into numeric format (0 for benign, 1 for malignant) [Sometimes numerical lables are more efficient in R language and we can see the use in the later confusion matrix].

The data is then transformed into an **XGBoost DMatrix**, an optimized format for efficient training.

Finally, an XGBoost model is trained using a **gradient boosting decision tree (GBDT)** with a maximum depth of 6, a learning rate (eta) of 0.5, and a binary logistic objective for classification.

```r
library(caret)
```

```
Loading required package: ggplot2
```

```
Loading required package: lattice
```

```r
library(xgboost)

features <- train[, -ncol(train)]
labels <- train[, 10]
labels_train <- as.numeric(labels) -1

dtrain <- xgb.DMatrix(data = as.matrix(features), label = labels_train)

model_xgb <- xgboost(data=dtrain,booster='gbtree',max_depth=6,eta=0.5,objective='binary:logis
```

```
[1] train-logloss:0.352832
[2] train-logloss:0.221138
[3] train-logloss:0.148604
[4] train-logloss:0.108287
[5] train-logloss:0.079035
[6] train-logloss:0.060746
[7] train-logloss:0.048573
[8] train-logloss:0.040847
[9] train-logloss:0.034820
[10]    train-logloss:0.030503
[11]    train-logloss:0.027346
[12]    train-logloss:0.025624
[13]    train-logloss:0.023835
[14]    train-logloss:0.022142
[15]    train-logloss:0.020450
[16]    train-logloss:0.019718
[17]    train-logloss:0.018847
[18]    train-logloss:0.018096
[19]    train-logloss:0.016893
[20]    train-logloss:0.016375
[21]    train-logloss:0.015945
[22]    train-logloss:0.015405
[23]    train-logloss:0.015048
[24]    train-logloss:0.014824
[25]    train-logloss:0.014117
```

We do the same extraction for the test dataset.

```
features_test <- test[, -ncol(train)]
labels_test <- test[, 10]

labels_test_num  <- as.numeric(labels_test) -1

dtest <- xgb.DMatrix(data = as.matrix(features_test), label = labels_test_num)
```

Here we perform the prediction using the trained XGBoost model and convert the results into binary class labels (0 vs. 1) according to their prrdiction probability.

```
prob <- predict(model_xgb , dtest)

prediction <- factor(prob > 0.6,levels=c(FALSE,TRUE),label=c("0","1"))

print(prediction)
```

```
  [1] 0 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 0 1 1 0 1 0 0 1 1 0 0 0 1 0 0 0 0 0 0 1 1
 [38] 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 0 0
 [75] 1 1 1 1 0 0 0 0 1 1 0 0 1 0 1 0 0 0 1 0 1 0 0 1 0 1 0 1 0 0 0 0 1 1 0 0 0
[112] 0 0 0 0 0 1 1 1 0 1 0 0 0 0 1 0 1 0 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0 0 1 0 1
[149] 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 1 0 0 0 0 0 1 1 0 1 1 1 0 0 0 0 0
[186] 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
Levels: 0 1
```

Now that we have obtained the predicted classification labels for the test data, we can evaluate the performance of our XGBoost model using a confusion matrix. This will allow us to compare the actual and predicted classifications to assess the model's accuracy.

This result automatically generates various evaluation metrics, such as sensitivity and accuracy. Unlike logistic regression and decision trees, which require manual calculation of these metrics, XGBoost provides them directly. Based on these metrics, our model performs quite well.

```
library(caret)
xgb.cf <-caret::confusionMatrix(as.factor(prediction),as.factor(labels_test_num))
xgb.cf
```

```
Confusion Matrix and Statistics

          Reference
```

```
Prediction   0   1
         0 139   4
         1   3  59

               Accuracy : 0.9659
                 95% CI : (0.9309, 0.9862)
    No Information Rate : 0.6927
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9194

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9789
            Specificity : 0.9365
         Pos Pred Value : 0.9720
         Neg Pred Value : 0.9516
             Prevalence : 0.6927
         Detection Rate : 0.6780
   Detection Prevalence : 0.6976
      Balanced Accuracy : 0.9577

       'Positive' Class : 0
```

Conclusion

XGBoost (Extreme Gradient Boosting) is an advanced ensemble learning technique based on decision trees. It improves model performance by iteratively optimizing weak learners and reducing errors using gradient boosting. Compared to traditional machine learning models, XGBoost is known for its high efficiency, scalability, and ability to handle complex patterns in data.

In theory, XGBoost performs best on large datasets with high-dimensional features and structured data. It is particularly effective when dealing with missing values and interactions between features. However, for our breast cancer classification task, all five models—Logistic Regression, Decision Trees, Random Forest, SVM, and XGBoost—have shown relatively strong performance. This suggests that the dataset is well-structured and suitable for various classification algorithms.