DM-HW3-Q2

a

1. 代码运行结果拷贝

1. User Features

781924	24	332.67	5	8.0	5.0	4.0	7.0	-0.4	101.65	88.52	62.0	80.5	-8.997000000000002	8.0
13325038116	17	1514.38	9	0.0	6.0	8.0	3.0	1.1	0.0	204.07999999999998	173.3	1137.0	338.02200000000005	0.0
13854627199	47	470.1300000000001	13	26.0	0.0	21.0	0.0	-5.7	251.54000000000005	0.0	218.59	0.0	-53.60300000000002	26.0
13864739266	39	299.36999999999995	4	0.0	13.0	16.0	10.0	3.3	0.0	110.4	99.42999999999998	89.54	25.764999999999997	0.0
15954611837	29	237.30000000000004	4	0.0	4.0	10.0	15.0	5.1	0.0	54.849999999999994	82.68	99.77000000000001	32.71400000000001	0.0
15954688237	42	499.81	5	0.0	0.0	19.0	23.0	8.8	0.0	0.0	152.38000000000002	347.43	119.46700000000001	0.0
15963883482	62	563.84	8	10.0	6.0	44.0	2.0	1.4	29.600000000000005	109.30999999999999	312.72999999999996	112.2	45.122	10.0
15963885355	25	174.78000000000003	4	3.0	11.0	11.0	0.0	-0.9	19.5	112.77	42.510000000000005	0.0	-12.876	3.0
18554652702	32	426.99	3	0.0	10.0	14.0	8.0	2.8	0.0	105.5	178.69000000000003	142.8	50.159000000000006	0.0
18654692914	43	360.71	7	9.0	9.0	21.0	4.0	-0.3	70.72	79.49	152.20000000000005	58.3	3.545000000000006	9.0
1590120464497	24	157.82	6	18.0	0.0	2.0	4.0	-4.0	132.72	0.0	4.7	20.4	-33.226000000000006	18.0
1590130640102	45	418.20000000000005	21	26.0	9.0	10.0	0.0	-7.7	133.0	225.82999999999998	59.370000000000005	0.0	-56.54600000000001	26.0
1590130817948	95	727.2600000000002	16	3.0	27.0	41.0	24.0	7.7	14.469999999999999	166.55000000000004	375.23	171.01000000000002	67.83	3.0
1590140304209	1	6.18	1	1.0	0.0	0.0	0.0	-0.3	6.18	0.0	0.0	0.0	-1.8539999999999999	1.0
1590140304506	84	939.9299999999996	15	14.0	9.0	26.0	35.0	8.0	129.76000000000002	141.5	303.14	365.5300000000001	86.89500000000002	14.0

.....

2. Item Features

10000000	1	10.8	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	10.8
10000003	1	5.9	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0
10000004	6	34.8	6	1.0	3.0	1.0	1.0	-0.2	5.8	17.4	5.8
10000005	4	8.0	4	0.0	2.0	1.0	1.0	0.2	0.0	4.0	2.0
10000006	20	54.00000000000002	15	10.0	4.0	3.0	3.0	-2.2	26.999999999999996	10.8	8.100000000000001
10000007	2	9.0	2	1.0	0.0	1.0	0.0	-0.2	4.5	0.0	4.5
10000009	5	32.5	4	0.0	2.0	3.0	0.0	0.1	0.0	13.0	19.5
10000010	7	43.400000000000006	6	0.0	2.0	1.0	4.0	1.1	0.0	12.4	6.2
10000011	16	43.20000000000001	12	5.0	4.0	4.0	3.0	-0.6	13.5	10.8	10.8
10000012	3	32.400000000000006	3	1.0	1.0	0.0	1.0	-0.1	10.8	10.8	0.0
10000013	3	6.6000000000000005	2	0.0	0.0	3.0	0.0	0.3	0.0	0.0	6.6000000000000005
10000015	7	18.9	6	6.0	0.0	0.0	1.0	-1.5	16.2	0.0	0.0

.....

3. Brand Features

-1.0	6732	71920.80000000051	122	1751.0	1790.0	1560.0	1631.0	-59.0	18105.609999999997	19099.329999999984	15600.83999999997
10003.0	3	20.5	3	0.0	1.0	1.0	1.0	0.3	0.0	3.5	5.0
10005.0	1	11.0	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	11.0
10007.0	41	139.29999999999995	28	14.0	10.0	8.0	9.0	-1.7	40.6	34.800000000000004	34.3
10012.0	6	326.0	3	1.0	4.0	0.0	1.0	-0.4	58.0	230.0	0.0
10015.0	3	25.0	3	0.0	0.0	2.0	1.0	0.5	0.0	0.0	21.5
10029.0	4	16.2	3	0.0	0.0	0.0	4.0	1.2	0.0	0.0	0.0
10033.0	25	271.0	15	11.0	0.0	6.0	8.0	-0.3	127.0	0.0	66.0
10036.0	1	10.0	1	0.0	1.0	0.0	0.0	-0.1	0.0	10.0	0.0
10038.0	1	16.2	1	0.0	1.0	0.0	0.0	-0.1	0.0	16.2	0.0
10045.0	24	197.10000000000005	16	2.0	9.0	3.0	10.0	1.8	6.0	57.5	35.6
10049.0	2	20.0	2	2.0	0.0	0.0	0.0	-0.6	20.0	0.0	0.0
10053.0	11	31.899999999999999	8	2.0	1.0	1.0	7.0	1.5	5.8	2.9	2.9
10054.0	1	158.0	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	158.0
10060.0	91	261.4600000000002	54	18.0	25.0	32.0	16.0	0.1	46.8	84.28	85.18000000000004
10062.0	45	193.19999999999993	31	13.0	9.0	13.0	10.0	-0.5	45.0	47.00000000000001	48.599999999999994
10063.0	4	9.8	3	0.0	1.0	3.0	0.0	0.2	0.0	2.7	7.1000000000000005

.....

4. Category Features

10000	90	344.6999999999998	59	25.0	18.0	24.0	23.0	0.0	80.80000000000003	79.2	93.9	9
10001	16	50.300000000000004	15	2.0	5.0	5.0	4.0	0.6	8.9	13.5	15.3	1
10002	61	389.2999999999999	41	15.0	17.0	19.0	10.0	-1.3	107.2	83.0	96.50000000000001	1
10008	17	90.39999999999999	16	6.0	6.0	5.0	0.0	-1.9	38.0	29.6	22.8	0
10100	16	191.40000000000003	13	0.0	6.0	5.0	5.0	1.4	0.0	32.3	82.0	7
10102	1	15.8	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	15.8	0
10103	4	36.2	4	0.0	1.0	0.0	3.0	0.8	0.0	3.5	0.0	3
10109	17	288.4000000000001	15	4.0	1.0	4.0	8.0	1.5	67.2	16.8	74.9	1
10110	27	116.49000000000002	22	3.0	8.0	8.0	8.0	1.5	13.9	27.89	32.7	4
10111	6	32.400000000000006	5	1.0	4.0	0.0	1.0	-0.4	6.8	21.6	0.0	4
10112	7	51.199999999999996	6	0.0	2.0	1.0	4.0	1.1	0.0	21.7	3.9	2
10113	10	47.49999999999999	10	0.0	4.0	5.0	1.0	0.4	0.0	19.5	20.099999999999998	7
10114	7	23.5	7	3.0	1.0	2.0	1.0	-0.5	11.5	2.5	7.0	2
10116	29	202.39999999999998	21	11.0	9.0	1.0	8.0	-1.7	84.19999999999999	68.5	4.5	4
10117	8	20.0	7	4.0	1.0	3.0	0.0	-1.0	10.0	2.5	7.5	0

.....

5. UI Features

781924	10113009	1	6.8	1	0.0	1.0	0.0	0.0	-0.1	0.0	6.8	0.0	0.0	-0.6799999999999999
781924	10130009	1	5.5	1	1.0	0.0	0.0	0.0	-0.3	5.5	0.0	0.0	0.0	-1.65
781924	11302032	1	10.8	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	10.8	3.2400000000000007
781924	11531020	1	15.6	1	1.0	0.0	0.0	0.0	-0.3	15.6	0.0	0.0	0.0	-4.68
781924	11532011	1	9.9	1	1.0	0.0	0.0	0.0	-0.3	9.9	0.0	0.0	0.0	-2.97
781924	11532036	1	3.5	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	3.5	0.0	0.35
781924	11533012	3	32.400000000000006	2	1.0	2.0	0.0	0.0	-0.5	10.8	21.6	0.0	0.0	-5.4
781924	14014034	1	10.9	1	1.0	0.0	0.0	0.0	-0.3	10.9	0.0	0.0	0.0	-3.2700000000000005
781924	14050019	1	5.9	1	1.0	0.0	0.0	0.0	-0.3	5.9	0.0	0.0	0.0	-1.7700000000000002
781924	14082002	1	3.9	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	3.9	1.17
781924	14101028	1	4.0	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	4.0	1.2
781924	14402009	1	15.45	1	1.0	0.0	0.0	0.0	-0.3	15.45	0.0	0.0	0.0	-4.635
781924	14403083	1	23.32	1	0.0	1.0	0.0	0.0	-0.1	0.0	23.32	0.0	0.0	-2.332
781924	15113000	1	16.8	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	16.8	0.0	1.6800000000000002
781924	15120000	1	15.0	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	15.0	4.5
781924	15200001	1	13.8	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	13.8	4.140000000000001
781924	15200007	3	101.2	3	1.0	1.0	1.0	0.0	-0.3	27.6	36.8	36.8	0.0	-8.280000000000001
781924	15202012	1	4.9	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	4.9	0.0	0.4900000000000005
781924	23113024	1	18.0	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	18.0	5.4

……

5. UB Features

781924	-1.0	5	73.72	3	1.0	3.0	0.0	1.0	-0.3	10.8	44.92	0.0	18.0	-2.3320000000000003
781924	10106.0	1	5.5	1	1.0	0.0	0.0	0.0	-0.3	5.5	0.0	0.0	0.0	-1.65
781924	10706.0	1	6.8	1	0.0	1.0	0.0	0.0	-0.1	0.0	6.8	0.0	0.0	-0.6799999999999999
781924	11129.0	1	9.9	1	1.0	0.0	0.0	0.0	-0.3	9.9	0.0	0.0	0.0	-2.97
781924	11149.0	1	15.6	1	1.0	0.0	0.0	0.0	-0.3	15.6	0.0	0.0	0.0	-4.68
781924	11154.0	1	3.5	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	3.5	0.0	0.35
781924	11288.0	1	10.8	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	10.8	3.2400000000000007
781924	14082.0	1	5.9	1	1.0	0.0	0.0	0.0	-0.3	5.9	0.0	0.0	0.0	-1.7700000000000002
781924	14350.0	1	15.45	1	1.0	0.0	0.0	0.0	-0.3	15.45	0.0	0.0	0.0	-4.635
781924	14436.0	1	4.0	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	4.0	1.2
781924	14731.0	1	3.9	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	3.9	1.17
781924	14759.0	1	10.9	1	1.0	0.0	0.0	0.0	-0.3	10.9	0.0	0.0	0.0	-3.2700000000000005
781924	15012.0	2	31.8	2	0.0	0.0	1.0	1.0	0.4	0.0	0.0	16.8	15.0	6.180000000000001
781924	15052.0	1	4.9	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	4.9	0.0	0.4900000000000005
781924	15094.0	3	101.2	3	1.0	1.0	1.0	0.0	-0.3	27.6	36.8	36.8	0.0	-8.280000000000001
781924	15631.0	1	13.8	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	13.8	4.140000000000001
781924	34001.0	1	15.0	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	15.0	4.5

……

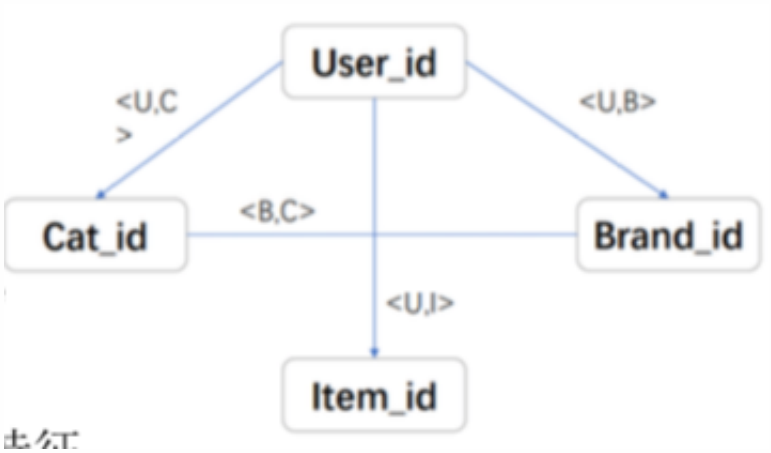
6. UC Features

781924	10113	1	6.8	1	0.0	1.0	0.0	0.0	-0.1	0.0	6.8	0.0	0.0	-0.6799999999999999
781924	10130	1	5.5	1	1.0	0.0	0.0	0.0	-0.3	5.5	0.0	0.0	0.0	-1.65
781924	11302	1	10.8	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	10.8	3.2400000000000007
781924	11531	1	15.6	1	1.0	0.0	0.0	0.0	-0.3	15.6	0.0	0.0	0.0	-4.68
781924	11532	2	13.4	2	1.0	0.0	1.0	0.0	-0.2	9.9	0.0	3.5	0.0	-2.62
781924	11533	3	32.400000000000006	2	1.0	2.0	0.0	0.0	-0.5	10.8	21.6	0.0	0.0	-5.4
781924	14014	1	10.9	1	1.0	0.0	0.0	0.0	-0.3	10.9	0.0	0.0	0.0	-3.2700000000000005
781924	14050	1	5.9	1	1.0	0.0	0.0	0.0	-0.3	5.9	0.0	0.0	0.0	-1.7700000000000002
781924	14082	1	3.9	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	3.9	1.17
781924	14101	1	4.0	1	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	4.0	1.2
781924	14402	1	15.45	1	1.0	0.0	0.0	0.0	-0.3	15.45	0.0	0.0	0.0	-4.635
781924	14403	1	23.32	1	0.0	1.0	0.0	0.0	-0.1	0.0	23.32	0.0	0.0	-2.332
781924	15113	1	16.8	1	0.0	0.0	1.0	0.0	0.1	0.0	0.0	16.8	0.0	1.6800000000000002

……

2. 讨论分析部分

1. 特征提取对象



1. 单个对象的特征：针对用户、品牌、类别、商品分别提取特征
2. 联合特征：用户-品牌之间的交互特征（用户-类别、用户-商品...）

2. 特征类型

- 特征类型：
  1. 整个时间段简单统计（购买次数、购买天数、消费金额...）
  2. 每月统计（同上）
  3. 聚合特征：按月统计的特征进行聚合...
  4. 最近时间段内的特征（最近一个月、最近一周）
  5. 复杂特征（用户购买的趋势、用户-品牌间的相似度）
  6. 时间特征（用户购买周期、最近一次购买到待预测时间段的距离...）

3. 性能比较图表

整个过程大概花10min左右的时间

b

1.代码运行结果拷贝

```
2900003115009, 22002240, NO
2900000476042, 15110032, NO
2900000476042, 15110032, NO
1590140304506, 27400855, NO
1591014637324, 15110032, NO
1591015420000, 27002269, NO
2900000386440, 15130013, YES
2900002932416, 15130013, NO
1590151103907, 14842024, NO
2900001234399, 15130013, NO
1591016517778, 11111001, NO
1590142201148, 14721034, NO
1590142201148, 14721034, NO
2900002944495, 22002242, NO
1592015058781, 14870003, NO
1590142203104, 27002262, NO
1592015058781, 14721034, NO
1591015454814, 14802020, NO
1594140460286, 22102012, NO
2900000284333, 22000031, NO
2900000386440, 22000031, NO
2900000386440, 22171005, NO
2900000345300, 22021260, NO
2900000345300, 22021260, NO
18554652702, 22007008, NO
2900002932416, 22007008, NO
2900000452756, 22021260, NO
2900000178106, 11300123, NO
```

2. 讨论分析部分

1. 选择User Features/Item Features/UI Features 的合集作为特征
2. 使用七个分类器进行分类
3. 根据auc和f-measurement，GaussianNB分类器的表现最好
4. 运行时间都较短且相差不大

3. 性能比较图表

分类器	GaussianNB	KNeighbors	DecisionTree	RandomForest	Bagging	AdaBoost	GradientBoosting
运行时间（s）	0.6534	0.6213	0.6976	0.6927	0.7477	0.7541	0.7663

ci

1. 代码运行结果拷贝

```
2900003115009, YES
2900000476042, NO
2900000476042, NO
1590140304506, YES
1591014637324, YES
1591015420000, YES
2900000386440, YES
2900002932416, YES
1590151103907, YES
2900001234399, YES
1591016517778, YES
1590142201148, NO
1590142201148, NO
2900002944495, YES
1592015058781, YES
1590142203104, YES
1592015058781, YES
1591015454814, YES
1594140460286, YES
2900000284333, YES
2900000386440, YES
2900000386440, YES
2900000345300, YES
2900000345300, YES
18554652702, YES
2900002932416, YES
2900000452756, YES
2900000178106, YES
2900002930092, YES
2900002930092, YES
1590140306678, YES
1591016440328, NO
```

2. 讨论分析部分

1. 选择User Features作为特征
2. 使用七个分类器进行分类

3. 性能比较图图表

分类器	GaussianNB	KNeighbors	DecisionTree	RandomForest	Bagging	AdaBoost	GradientBoosting
运行时间（s）	0.3947	0.3566	0.4269	0.4387	0.4845	0.4962	0.5172

cii

1. 代码运行结果拷贝

```
2900003115009, -1.0, NO
2900000476042, 15039.0, NO
2900000476042, 15039.0, NO
1590140304506, -1.0, YES
1591014637324, 15039.0, NO
1591015420000, -1.0, YES
2900000386440, 15039.0, YES
2900002932416, 15039.0, NO
1590151103907, 14330.0, NO
2900001234399, 15039.0, NO
1591016517778, -1.0, YES
1590142201148, 14177.0, NO
1590142201148, 14177.0, NO
2900002944495, -1.0, NO
1592015058781, 14126.0, NO
1590142203104, -1.0, YES
1592015058781, 14177.0, NO
1591015454814, 14200.0, YES
1594140460286, -1.0, YES
2900000284333, -1.0, NO
2900000386440, -1.0, YES
2900000386440, -1.0, YES
2900000345300, -1.0, YES
2900000345300, -1.0, YES
18554652702, -1.0, YES
2900002932416, -1.0, YES
2900000452756, -1.0, NO
2900000178106, 11114.0, NO
2900002930092, 11150.0, NO
2900002930092, 11150.0, NO
```

2. 讨论分析部分

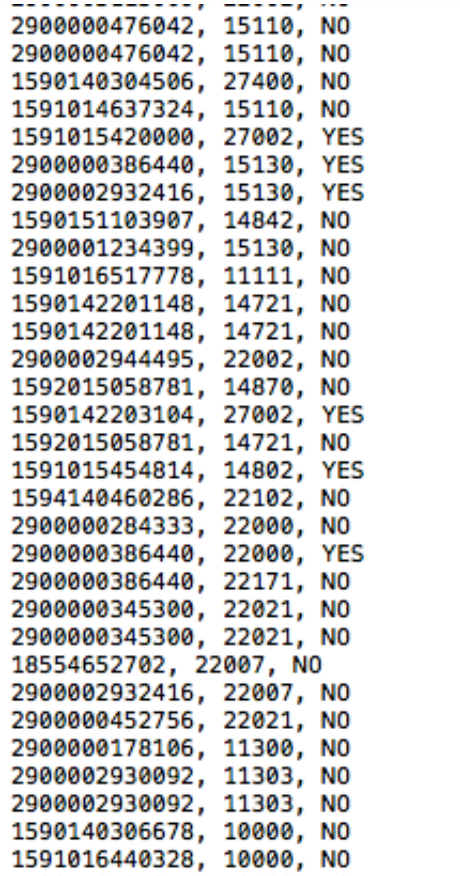
1. 选择User Features/Brand Features/U Features作为特征
2. 使用七个分类器进行分类

3. 性能比较图图表

分类器	GaussianNB	KNeighbors	DecisionTree	RandomForest	Bagging	AdaBoost	GradientBoosting
运行时间（s）	0.4098	0.3972	0.4683	0.4761	0.4892	0.4961	0.5211

ciii

1. 运行结果拷贝



2. 讨论分析部分

1. 选择User Features/Category Features/UC Features作为特征

2. 使用七个分类器进行分类

3. 性能比较图图表

分类器	GaussianNB	KNeighbors	DecisionTree	RandomForest	Bagging	AdaBoost	GradientBoosting
运行时间（s）	0.4248	0.3612	0.4387	0.4416	0.4625	0.4762	0.5046

civ

1. 运行结果拷贝

```
2900003115009, 5.3
1591016411137, 38.3
2900000448520, 370.1
2900001540117, 18.1
1591015184407, 10.6
18554652702, 0.7
1590142648356, 64.8
1594140121125, 49.8
1590142205993, 142.2
1591015034924, 9.3
2900000124974, 174.3
1598140129341, 40.1
2900000071742, 66.3
1593160622155, 4.1
2900001509459, 107.7
2900000401495, 6.2
1595151738968, 44.7
2900001364058, 92.3
1591015454814, 87.6
1591015420000, 55.5
2900001431651, 0.4
2900000680036, 9.2
2900001216616, 106.1
2900000557161, 5.5
1590142191722, 104.5
2900002936940, 118.4
2900002934892, 123.9
1595151630446, 59.3
1590120464497, 67.1
```

2. 讨论分析部分

1. 选择User Features作为特征
2. 使用七个回归模型进行回归

3. 性能比较图图表

分类器	GaussianNB	KNeighbors	DecisionTree	RandomForest	Bagging	AdaBoost	GradientBoosting
运行时间（s）	0.4689	0.4373	0.4669	0.4751	0.4727	0.4834	0.4936