forest **=** RandomForestClassifier**(**n\_estimators**=**10000, oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 4.您平时参加体育锻炼的频率是: \* 0.048337

2) 10(计步/计数) 0.048134

3) 10(视频指导/运动知识) 0.043535

4) 9 (B.空闲时间) 0.040178

5) 7.您偏向的运动环境是: \* 0.037676

6) 10(卡路里消耗) 0.035026

7) 10(GPS记录运动轨迹) 0.032601

8) 10(运动计划) 0.031052

9) 10(音乐播放) 0.030910

10) 计步/计数获知日常运动量 0.028143

11) APP运动达人及好友的经验分享 0.027868

12) 运动排行榜或好友间竞赛的激励方式 0.026509

13) 9 (C.场地器材) 0.026325

14) 运动知识、视频指导的全面性与专业性程度 0.026291

15) 10(运动圈互动) 0.026047

16) 9 (F.集体效应) 0.025657

17) 2.您所在的年级是: \* 0.024709

18) 3.您更加喜欢的颜色系是: \* 0.023639

19) 5.您每次进行体育锻炼的时间是: \* 0.023121

20) 使用运动类app的性价比优势（对比健身房等） 0.023050

21) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.022954

22) 9 (A.锻炼兴趣) 0.022844

23) 6.您每次进行体育锻炼的强度是: \* 0.022838

24) 计算卡路里消耗 0.022756

25) 9 (D.专业辅导) 0.022354

26) 10(提供课程性价比高) 0.021626

27) 音乐播放愉悦运动心情 0.021328

28) 10(运动榜单/好友竞赛) 0.021106

29) 制定运动计划和运动安排的执行程度 0.020751

30) 1.您的性别是: \* 0.020736

31) 9 (E.媒体宣传) 0.018770

32) 10(其他) 0.016497

33) 8 (E.排球) 0.014904

34) 8 (J.瑜伽) 0.014282

35) 9 (G.其他) 0.013479

36) 8 (A.跑步) 0.011454

37) 8 (C.羽毛球) 0.011129

38) 8 (B.篮球) 0.010031

39) 8 (D.兵乓球) 0.007870

40) 8 (I.跳绳) 0.006958

41) 8 (H.健美操) 0.006330

42) 8 (G.足球) 0.005831

43) 8 (K.游泳) 0.005672

44) 8 (其他) 0.005623

45) 8 (F.网球) 0.003069

0.7641509433962265

AUC Score (Train): 1.000000

进程已结束，退出代码 0

forest **=** RandomForestClassifier**(**n\_estimators**=**529, oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 10(计步/计数) 0.049480

2) 4.您平时参加体育锻炼的频率是: \* 0.045896

3) 10(视频指导/运动知识) 0.045506

4) 9 (B.空闲时间) 0.042202

5) 7.您偏向的运动环境是: \* 0.035711

6) 10(GPS记录运动轨迹) 0.033417

7) 10(运动计划) 0.032788

8) 10(音乐播放) 0.032762

9) 10(卡路里消耗) 0.032537

10) APP运动达人及好友的经验分享 0.028493

11) 运动排行榜或好友间竞赛的激励方式 0.028303

12) 计步/计数获知日常运动量 0.027632

13) 运动知识、视频指导的全面性与专业性程度 0.026455

14) 9 (C.场地器材) 0.026039

15) 9 (F.集体效应) 0.025748

16) 6.您每次进行体育锻炼的强度是: \* 0.025002

17) 5.您每次进行体育锻炼的时间是: \* 0.024876

18) 10(运动圈互动) 0.024736

19) 2.您所在的年级是: \* 0.024091

20) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.023990

21) 使用运动类app的性价比优势（对比健身房等） 0.022587

22) 9 (A.锻炼兴趣) 0.022545

23) 计算卡路里消耗 0.022239

24) 9 (D.专业辅导) 0.022194

25) 音乐播放愉悦运动心情 0.021856

26) 10(运动榜单/好友竞赛) 0.021471

27) 10(提供课程性价比高) 0.021252

28) 3.您更加喜欢的颜色系是: \* 0.020765

29) 1.您的性别是: \* 0.019786

30) 制定运动计划和运动安排的执行程度 0.018880

31) 9 (E.媒体宣传) 0.017855

32) 10(其他) 0.016585

33) 8 (E.排球) 0.014549

34) 8 (J.瑜伽) 0.014238

35) 9 (G.其他) 0.014088

36) 8 (C.羽毛球) 0.011563

37) 8 (A.跑步) 0.011189

38) 8 (B.篮球) 0.009745

39) 8 (D.兵乓球) 0.007849

40) 8 (I.跳绳) 0.006608

41) 8 (H.健美操) 0.006402

42) 8 (G.足球) 0.006099

43) 8 (其他) 0.005300

44) 8 (K.游泳) 0.005247

45) 8 (F.网球) 0.003444

0.7688679245283019

AUC Score (Train): 1.000000

进程已结束，退出代码 0

forest **=** RandomForestClassifier**(**n\_estimators**=**529, oob\_score**=True**,min\_samples\_split**=**110,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 4.您平时参加体育锻炼的频率是: \* 0.078429

2) 9 (B.空闲时间) 0.074195

3) 10(其他) 0.071261

4) 10(视频指导/运动知识) 0.067342

5) 7.您偏向的运动环境是: \* 0.064264

6) 1.您的性别是: \* 0.045214

7) 10(运动圈互动) 0.041631

8) 8 (E.排球) 0.041139

9) 计步/计数获知日常运动量 0.037209

10) 9 (G.其他) 0.029243

11) 运动知识、视频指导的全面性与专业性程度 0.029079

12) 运动排行榜或好友间竞赛的激励方式 0.028492

13) 10(提供课程性价比高) 0.026980

14) 8 (J.瑜伽) 0.024786

15) 3.您更加喜欢的颜色系是: \* 0.023874

16) 2.您所在的年级是: \* 0.023254

17) APP运动达人及好友的经验分享 0.023069

18) 8 (B.篮球) 0.022990

19) 计算卡路里消耗 0.022926

20) 5.您每次进行体育锻炼的时间是: \* 0.021645

21) 10(计步/计数) 0.019076

22) 10(GPS记录运动轨迹) 0.017352

23) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.017249

24) 9 (F.集体效应) 0.015986

25) 制定运动计划和运动安排的执行程度 0.014294

26) 10(卡路里消耗) 0.013952

27) 10(音乐播放) 0.013806

28) 9 (A.锻炼兴趣) 0.012074

29) 使用运动类app的性价比优势（对比健身房等） 0.011466

30) 音乐播放愉悦运动心情 0.009625

31) 6.您每次进行体育锻炼的强度是: \* 0.009544

32) 8 (C.羽毛球) 0.008396

33) 10(运动计划) 0.008188

34) 8 (A.跑步) 0.006208

35) 10(运动榜单/好友竞赛) 0.005386

36) 9 (D.专业辅导) 0.004538

37) 9 (C.场地器材) 0.004080

38) 8 (I.跳绳) 0.003237

39) 9 (E.媒体宣传) 0.002711

40) 8 (F.网球) 0.002241

41) 8 (其他) 0.001370

42) 8 (D.兵乓球) 0.001305

43) 8 (G.足球) 0.000539

44) 8 (H.健美操) 0.000359

45) 8 (K.游泳) 0.000000

0.7783018867924528

AUC Score (Train): 0.851709

进程已结束，退出代码 0

forest **=** RandomForestClassifier**(**n\_estimators**=**529,min\_samples\_split**=**90,oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

forest **=** RandomForestClassifier**(**n\_estimators**=**529,min\_samples\_split**=**90,max\_depth**=**5,oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 9 (B.空闲时间) 0.082081

2) 4.您平时参加体育锻炼的频率是: \* 0.080386

3) 7.您偏向的运动环境是: \* 0.063229

4) 10(其他) 0.059053

5) 10(视频指导/运动知识) 0.057747

6) 1.您的性别是: \* 0.054011

7) 8 (E.排球) 0.042447

8) 计算卡路里消耗 0.030820

9) 计步/计数获知日常运动量 0.030795

10) 运动排行榜或好友间竞赛的激励方式 0.030011

11) 10(运动圈互动) 0.029157

12) 10(提供课程性价比高) 0.028830

13) 8 (J.瑜伽) 0.026955

14) APP运动达人及好友的经验分享 0.026394

15) 9 (G.其他) 0.024616

16) 3.您更加喜欢的颜色系是: \* 0.023598

17) 5.您每次进行体育锻炼的时间是: \* 0.023289

18) 运动知识、视频指导的全面性与专业性程度 0.022619

19) 9 (F.集体效应) 0.021517

20) 10(GPS记录运动轨迹) 0.021368

21) 10(计步/计数) 0.019552

22) 2.您所在的年级是: \* 0.019076

23) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.015545

24) 10(音乐播放) 0.015316

25) 6.您每次进行体育锻炼的强度是: \* 0.014795

26) 使用运动类app的性价比优势（对比健身房等） 0.014751

27) 9 (A.锻炼兴趣) 0.013617

28) 音乐播放愉悦运动心情 0.012396

29) 8 (B.篮球) 0.012142

30) 10(卡路里消耗) 0.011993

31) 制定运动计划和运动安排的执行程度 0.011811

32) 10(运动计划) 0.010150

33) 9 (C.场地器材) 0.007784

34) 8 (C.羽毛球) 0.007610

35) 10(运动榜单/好友竞赛) 0.007011

36) 9 (D.专业辅导) 0.006328

37) 8 (A.跑步) 0.005562

38) 8 (I.跳绳) 0.003881

39) 9 (E.媒体宣传) 0.003308

40) 8 (F.网球) 0.003235

41) 8 (D.兵乓球) 0.001769

42) 8 (其他) 0.001269

43) 8 (K.游泳) 0.000920

44) 8 (H.健美操) 0.000748

45) 8 (G.足球) 0.000510

0.7783018867924528

AUC Score (Train): 0.865764

进程已结束，退出代码 0

forest **=** RandomForestClassifier**(**n\_estimators**=**529,min\_samples\_split**=**108,oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 4.您平时参加体育锻炼的频率是: \* 0.077487

2) 9 (B.空闲时间) 0.077208

3) 10(其他) 0.071917

4) 10(视频指导/运动知识) 0.063819

5) 7.您偏向的运动环境是: \* 0.062349

6) 1.您的性别是: \* 0.045214

7) 8 (E.排球) 0.041174

8) 10(运动圈互动) 0.036873

9) 计步/计数获知日常运动量 0.035460

10) 运动知识、视频指导的全面性与专业性程度 0.030370

11) 9 (G.其他) 0.029984

12) 运动排行榜或好友间竞赛的激励方式 0.028492

13) 10(提供课程性价比高) 0.026980

14) 8 (J.瑜伽) 0.026030

15) APP运动达人及好友的经验分享 0.024120

16) 3.您更加喜欢的颜色系是: \* 0.023995

17) 计算卡路里消耗 0.023993

18) 8 (B.篮球) 0.022990

19) 2.您所在的年级是: \* 0.022370

20) 5.您每次进行体育锻炼的时间是: \* 0.021645

21) 10(计步/计数) 0.018564

22) 10(GPS记录运动轨迹) 0.017743

23) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.016809

24) 9 (F.集体效应) 0.015986

25) 10(卡路里消耗) 0.014896

26) 10(音乐播放) 0.014632

27) 制定运动计划和运动安排的执行程度 0.014425

28) 使用运动类app的性价比优势（对比健身房等） 0.011595

29) 9 (A.锻炼兴趣) 0.011388

30) 6.您每次进行体育锻炼的强度是: \* 0.010664

31) 音乐播放愉悦运动心情 0.009466

32) 8 (C.羽毛球) 0.009145

33) 10(运动计划) 0.008188

34) 8 (A.跑步) 0.006208

35) 10(运动榜单/好友竞赛) 0.005659

36) 9 (D.专业辅导) 0.005589

37) 9 (C.场地器材) 0.004080

38) 8 (I.跳绳) 0.003237

39) 9 (E.媒体宣传) 0.002711

40) 8 (F.网球) 0.002241

41) 8 (D.兵乓球) 0.002039

42) 8 (其他) 0.001370

43) 8 (G.足球) 0.000539

44) 8 (H.健美操) 0.000359

45) 8 (K.游泳) 0.000000

0.7783018867924528

AUC Score (Train): 0.850935

进程已结束，退出代码 0

forest **=** RandomForestClassifier**(**n\_estimators**=**529,min\_samples\_split**=**108,min\_samples\_leaf**=**14,oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 9 (B.空闲时间) 0.088747

2) 10(其他) 0.075892

3) 10(视频指导/运动知识) 0.061225

4) 7.您偏向的运动环境是: \* 0.053724

5) 4.您平时参加体育锻炼的频率是: \* 0.053198

6) 1.您的性别是: \* 0.044972

7) 10(运动圈互动) 0.042250

8) 8 (E.排球) 0.040969

9) 9 (G.其他) 0.036527

10) 10(提供课程性价比高) 0.036192

11) 3.您更加喜欢的颜色系是: \* 0.032861

12) 计步/计数获知日常运动量 0.032548

13) 运动知识、视频指导的全面性与专业性程度 0.031793

14) 运动排行榜或好友间竞赛的激励方式 0.031138

15) 8 (B.篮球) 0.025906

16) 计算卡路里消耗 0.023436

17) 2.您所在的年级是: \* 0.023177

18) 制定运动计划和运动安排的执行程度 0.021525

19) 9 (A.锻炼兴趣) 0.019877

20) 9 (F.集体效应) 0.019299

21) 10(GPS记录运动轨迹) 0.018739

22) 8 (J.瑜伽) 0.018229

23) 10(计步/计数) 0.018216

24) 10(卡路里消耗) 0.016773

25) 5.您每次进行体育锻炼的时间是: \* 0.016315

26) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.014584

27) 10(音乐播放) 0.013842

28) 音乐播放愉悦运动心情 0.012154

29) 6.您每次进行体育锻炼的强度是: \* 0.008933

30) 10(运动计划) 0.008886

31) 8 (C.羽毛球) 0.008110

32) 使用运动类app的性价比优势（对比健身房等） 0.007609

33) 9 (C.场地器材) 0.006726

34) APP运动达人及好友的经验分享 0.006277

35) 8 (A.跑步) 0.006208

36) 10(运动榜单/好友竞赛) 0.005345

37) 9 (D.专业辅导) 0.005095

38) 9 (E.媒体宣传) 0.004567

39) 8 (D.兵乓球) 0.002482

40) 8 (I.跳绳) 0.002051

41) 8 (其他) 0.001979

42) 8 (F.网球) 0.001265

43) 8 (H.健美操) 0.000359

44) 8 (G.足球) 0.000000

45) 8 (K.游泳) 0.000000

0.7783018867924528

AUC Score (Train): 0.823598

进程已结束，退出代码 0

forest **=** RandomForestClassifier**(**n\_estimators**=**529,min\_samples\_split**=**88,oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 9 (B.空闲时间) 0.085594

2) 4.您平时参加体育锻炼的频率是: \* 0.078670

3) 7.您偏向的运动环境是: \* 0.063165

4) 10(视频指导/运动知识) 0.057633

5) 1.您的性别是: \* 0.055447

6) 10(其他) 0.052817

7) 8 (E.排球) 0.041039

8) 计算卡路里消耗 0.032376

9) 计步/计数获知日常运动量 0.030519

10) 运动排行榜或好友间竞赛的激励方式 0.029374

11) 10(提供课程性价比高) 0.028602

12) 10(运动圈互动) 0.028562

13) APP运动达人及好友的经验分享 0.028055

14) 9 (G.其他) 0.026661

15) 8 (J.瑜伽) 0.025458

16) 3.您更加喜欢的颜色系是: \* 0.023598

17) 5.您每次进行体育锻炼的时间是: \* 0.023102

18) 运动知识、视频指导的全面性与专业性程度 0.022619

19) 10(GPS记录运动轨迹) 0.022209

20) 9 (F.集体效应) 0.021517

21) 2.您所在的年级是: \* 0.018986

22) 10(计步/计数) 0.018555

23) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.016067

24) 10(音乐播放) 0.015032

25) 9 (A.锻炼兴趣) 0.014643

26) 使用运动类app的性价比优势（对比健身房等） 0.014219

27) 6.您每次进行体育锻炼的强度是: \* 0.012932

28) 10(卡路里消耗) 0.012279

29) 8 (B.篮球) 0.012110

30) 音乐播放愉悦运动心情 0.011885

31) 制定运动计划和运动安排的执行程度 0.010830

32) 10(运动计划) 0.010150

33) 9 (C.场地器材) 0.008346

34) 10(运动榜单/好友竞赛) 0.007843

35) 8 (C.羽毛球) 0.007610

36) 9 (E.媒体宣传) 0.006487

37) 9 (D.专业辅导) 0.005714

38) 8 (A.跑步) 0.005096

39) 8 (I.跳绳) 0.004476

40) 8 (F.网球) 0.003235

41) 8 (其他) 0.002019

42) 8 (D.兵乓球) 0.001769

43) 8 (H.健美操) 0.001192

44) 8 (K.游泳) 0.000999

45) 8 (G.足球) 0.000510

0.7783018867924528

AUC Score (Train): 0.870019

进程已结束，退出代码 0

# 最终

Code:

#!/usr/bin/env python  
# -\*- coding: utf-8 -\*-  
# @Time : 2019/6/17 11:40  
# @Author : Poet Yin  
# @Email : poet@poetyin.com  
# @Site :   
# @File : RFtolearn.py  
# @Software: PyCharm  
**import** pandas **as** pd  
**import** numpy **as** np  
**from** sklearn.model\_selection **import** train\_test\_split  
**from** sklearn.ensemble **import** RandomForestClassifier  
**from** sklearn **import** metrics  
df **=** pd.read\_csv**("dodata\_303.csv")**x, y **=** df.iloc**[:**, 1**:]**.values, df.iloc**[:**, 0**]**.values  
x\_train, x\_test, y\_train, y\_test **=** train\_test\_split**(**x, y, test\_size **=** 0.3, random\_state **=** 0**)**feat\_labels **=** df.columns**[**1**:]**forest **=** RandomForestClassifier**(**n\_estimators**=**529,min\_samples\_split**=**88,max\_depth**=**5,oob\_score**=True**,random\_state**=**0, n\_jobs**=-**1**)**forest.fit**(**x\_train, y\_train**)**importances **=** forest.feature\_importances\_  
indices **=** np.argsort**(**importances**)[::-**1**]  
for** f **in** range**(**x\_train.shape**[**1**]):** print**("%2d) %-\*s %f" % (**f **+** 1, 30, feat\_labels**[**indices**[**f**]]**, importances**[**indices**[**f**]]))**print**('\n')**print **(**forest.oob\_score\_**)**y\_predprob **=** forest.predict\_proba**(**x\_train**)[:**,1**]**print **("AUC Score (Train): %f" %** metrics.roc\_auc\_score**(**y\_train, y\_predprob**))**print**("RF training set score: %f" %** forest.score**(**x\_train, y\_train**))**print**("RF test set score: %f" %** forest.score**(**x\_test, y\_test**))**

E:\ProgramData\Python\python.exe "G:/学习/数学建模 实战/2019统计建模/data/RFtolearn.py"

1) 9 (B.空闲时间) 0.085594

2) 4.您平时参加体育锻炼的频率是: \* 0.078670

3) 7.您偏向的运动环境是: \* 0.063165

4) 10(视频指导/运动知识) 0.057633

5) 1.您的性别是: \* 0.055447

6) 10(其他) 0.052817

7) 8 (E.排球) 0.041039

8) 计算卡路里消耗 0.032376

9) 计步/计数获知日常运动量 0.030519

10) 运动排行榜或好友间竞赛的激励方式 0.029374

11) 10(提供课程性价比高) 0.028602

12) 10(运动圈互动) 0.028562

13) APP运动达人及好友的经验分享 0.028055

14) 9 (G.其他) 0.026661

15) 8 (J.瑜伽) 0.025458

16) 3.您更加喜欢的颜色系是: \* 0.023598

17) 5.您每次进行体育锻炼的时间是: \* 0.023102

18) 运动知识、视频指导的全面性与专业性程度 0.022619

19) 10(GPS记录运动轨迹) 0.022209

20) 9 (F.集体效应) 0.021517

21) 2.您所在的年级是: \* 0.018986

22) 10(计步/计数) 0.018555

23) 11.您对运动类APP在以下属性方面的重要性程度的看法是: \*—GPS运动轨迹记录运动历程 0.016067

24) 10(音乐播放) 0.015032

25) 9 (A.锻炼兴趣) 0.014643

26) 使用运动类app的性价比优势（对比健身房等） 0.014219

27) 6.您每次进行体育锻炼的强度是: \* 0.012932

28) 10(卡路里消耗) 0.012279

29) 8 (B.篮球) 0.012110

30) 音乐播放愉悦运动心情 0.011885

31) 制定运动计划和运动安排的执行程度 0.010830

32) 10(运动计划) 0.010150

33) 9 (C.场地器材) 0.008346

34) 10(运动榜单/好友竞赛) 0.007843

35) 8 (C.羽毛球) 0.007610

36) 9 (E.媒体宣传) 0.006487

37) 9 (D.专业辅导) 0.005714

38) 8 (A.跑步) 0.005096

39) 8 (I.跳绳) 0.004476

40) 8 (F.网球) 0.003235

41) 8 (其他) 0.002019

42) 8 (D.兵乓球) 0.001769

43) 8 (H.健美操) 0.001192

44) 8 (K.游泳) 0.000999

45) 8 (G.足球) 0.000510

0.7783018867924528

AUC Score (Train): 0.870019

RF training set score: 0.778302

RF test set score: 0.824176

进程已结束，退出代码 0