Kaggle_Scikit01

November 19, 2016

```
In [1]: import pandas as pd

In [2]: train = pd.read_csv('http://bit.ly/kaggletrain')

In [4]: train.head(2)

Out[4]:    PassengerId  Survived  Pclass  \
        0            1         0       3
        1            2         1       1

                                                        Name     Sex   Age  SibSp
        0                         Braund, Mr. Owen Harris    male  22.0      1
        1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1

           Parch      Ticket      Fare Cabin Embarked
        0      0   A/5 21171    7.2500   NaN        S
        1      0    PC 17599   71.2833   C85        C

In [5]: feature_cols = ['Pclass', 'Parch']

In [7]: X = train.loc[:, feature_cols]

In [8]: X.shape

Out[8]: (891, 2)

In [9]: y = train.Survived

In [10]: y.shape

Out[10]: (891,)

In [11]: from sklearn.linear_model import LogisticRegression
         logreg = LogisticRegression()
         logreg.fit(X,y)

Out[11]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=Tru
                   intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                   penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
                   verbose=0, warm_start=False)
```

```
In [12]: test = pd.read_csv('http://bit.ly/kaggletest')

In [13]: test.head()

Out[13]:    PassengerId  Pclass                                          Name    S
         0          892       3                              Kelly, Mr. James    ma
         1          893       3              Wilkes, Mrs. James (Ellen Needs)  fema
         2          894       2                     Myles, Mr. Thomas Francis    ma
         3          895       3                              Wirz, Mr. Albert    ma
         4          896       3  Hirvonen, Mrs. Alexander (Helga E Lindqvist)  fema

             Age  SibSp  Parch    Ticket     Fare Cabin Embarked
         0  34.5      0      0    330911   7.8292   NaN        Q
         1  47.0      1      0    363272   7.0000   NaN        S
         2  62.0      0      0    240276   9.6875   NaN        Q
         3  27.0      0      0    315154   8.6625   NaN        S
         4  22.0      1      1   3101298  12.2875   NaN        S

In [14]: X_new = test.loc[:, feature_cols]

In [15]: X_new.shape

Out[15]: (418, 2)

In [16]: new_pred_class = logreg.predict(X_new)

In [17]: test.PassengerId

Out[17]: 0      892
         1      893
         2      894
         3      895
         4      896
         5      897
         6      898
         7      899
         8      900
         9      901
         10     902
         11     903
         12     904
         13     905
         14     906
         15     907
         16     908
         17     909
         18     910
         19     911
         20     912
```

```
        21        913
        22        914
        23        915
        24        916
        25        917
        26        918
        27        919
        28        920
        29        921
                  ...
        388       1280
        389       1281
        390       1282
        391       1283
        392       1284
        393       1285
        394       1286
        395       1287
        396       1288
        397       1289
        398       1290
        399       1291
        400       1292
        401       1293
        402       1294
        403       1295
        404       1296
        405       1297
        406       1298
        407       1299
        408       1300
        409       1301
        410       1302
        411       1303
        412       1304
        413       1305
        414       1306
        415       1307
        416       1308
        417       1309
        Name: PassengerId, dtype: int64

In [18]: new_pred_class

Out[18]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1
                1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0
                1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1
                1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0
```

```
               1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1
               0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0
               0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0
               0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0
               1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0
               0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0
               0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1
               0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0
               0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1
               0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0
               0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0
               0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0
               1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1
               1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0
               1, 0, 0, 0])
```

```
In [21]: pd.DataFrame({'PassengerId' : test.PassengerId, 'Survived' : new_pred_clas

In [22]: train.to_pickle('train.pkl')

In [23]: pd.read_pickle('train.pkl')
```

```
Out[23]:      PassengerId  Survived  Pclass  \
         0              1         0       3
         1              2         1       1
         2              3         1       3
         3              4         1       1
         4              5         0       3
         5              6         0       3
         6              7         0       1
         7              8         0       3
         8              9         1       3
         9             10         1       2
         10            11         1       3
         11            12         1       1
         12            13         0       3
         13            14         0       3
         14            15         0       3
         15            16         1       2
         16            17         0       3
         17            18         1       2
         18            19         0       3
         19            20         1       3
         20            21         0       2
         21            22         1       2
         22            23         1       3
         23            24         1       1
         24            25         0       3
         25            26         1       3
```

4

26	27	0	3
27	28	0	1
28	29	1	3
29	30	0	3
..	...	...	...
861	862	0	2
862	863	1	1
863	864	0	3
864	865	0	2
865	866	1	2
866	867	1	2
867	868	0	1
868	869	0	3
869	870	1	3
870	871	0	3
871	872	1	1
872	873	0	1
873	874	0	3
874	875	1	2
875	876	1	3
876	877	0	3
877	878	0	3
878	879	0	3
879	880	1	1
880	881	1	2
881	882	0	3
882	883	0	3
883	884	0	2
884	885	0	3
885	886	0	3
886	887	0	2
887	888	1	1
888	889	0	3
889	890	1	1
890	891	0	3

	Name	Sex	Age	SibS
0	Braund, Mr. Owen Harris	male	22.0	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	
2	Heikkinen, Miss. Laina	female	26.0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	
4	Allen, Mr. William Henry	male	35.0	
5	Moran, Mr. James	male	NaN	
6	McCarthy, Mr. Timothy J	male	54.0	
7	Palsson, Master. Gosta Leonard	male	2.0	
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	
10	Sandstrom, Miss. Marguerite Rut	female	4.0	

11	Bonnell, Miss. Elizabeth	female	58.0
12	Saundercock, Mr. William Henry	male	20.0
13	Andersson, Mr. Anders Johan	male	39.0
14	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0
15	Hewlett, Mrs. (Mary D Kingcome)	female	55.0
16	Rice, Master. Eugene	male	2.0
17	Williams, Mr. Charles Eugene	male	NaN
18	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0
19	Masselmani, Mrs. Fatima	female	NaN
20	Fynney, Mr. Joseph J	male	35.0
21	Beesley, Mr. Lawrence	male	34.0
22	McGowan, Miss. Anna "Annie"	female	15.0
23	Sloper, Mr. William Thompson	male	28.0
24	Palsson, Miss. Torborg Danira	female	8.0
25	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38.0
26	Emir, Mr. Farred Chehab	male	NaN
27	Fortune, Mr. Charles Alexander	male	19.0
28	O'Dwyer, Miss. Ellen "Nellie"	female	NaN
29	Todoroff, Mr. Lalio	male	NaN
..	...	...	...
861	Giles, Mr. Frederick Edward	male	21.0
862	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48.0
863	Sage, Miss. Dorothy Edith "Dolly"	female	NaN
864	Gill, Mr. John William	male	24.0
865	Bystrom, Mrs. (Karolina)	female	42.0
866	Duran y More, Miss. Asuncion	female	27.0
867	Roebling, Mr. Washington Augustus II	male	31.0
868	van Melkebeke, Mr. Philemon	male	NaN
869	Johnson, Master. Harold Theodor	male	4.0
870	Balkic, Mr. Cerin	male	26.0
871	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0
872	Carlsson, Mr. Frans Olof	male	33.0
873	Vander Cruyssen, Mr. Victor	male	47.0
874	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0
875	Najib, Miss. Adele Kiamie "Jane"	female	15.0
876	Gustafsson, Mr. Alfred Ossian	male	20.0
877	Petroff, Mr. Nedelio	male	19.0
878	Laleff, Mr. Kristo	male	NaN
879	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0
880	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0
881	Markun, Mr. Johann	male	33.0
882	Dahlberg, Miss. Gerda Ulrika	female	22.0
883	Banfield, Mr. Frederick James	male	28.0
884	Sutehall, Mr. Henry Jr	male	25.0
885	Rice, Mrs. William (Margaret Norton)	female	39.0
886	Montvila, Rev. Juozas	male	27.0
887	Graham, Miss. Margaret Edith	female	19.0
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN

```
889                                       Behr, Mr. Karl Howell   male  26.0
890                                       Dooley, Mr. Patrick     male  32.0

     Parch            Ticket        Fare        Cabin  Embarked
0        0         A/5 21171      7.2500          NaN         S
1        0          PC 17599     71.2833          C85         C
2        0    STON/O2. 3101282    7.9250          NaN         S
3        0            113803     53.1000         C123         S
4        0            373450      8.0500          NaN         S
5        0            330877      8.4583          NaN         Q
6        0             17463     51.8625          E46         S
7        1            349909     21.0750          NaN         S
8        2            347742     11.1333          NaN         S
9        0            237736     30.0708          NaN         C
10       1           PP 9549     16.7000           G6         S
11       0            113783     26.5500         C103         S
12       0         A/5. 2151      8.0500          NaN         S
13       5            347082     31.2750          NaN         S
14       0            350406      7.8542          NaN         S
15       0            248706     16.0000          NaN         S
16       1            382652     29.1250          NaN         Q
17       0            244373     13.0000          NaN         S
18       0            345763     18.0000          NaN         S
19       0              2649      7.2250          NaN         C
20       0            239865     26.0000          NaN         S
21       0            248698     13.0000          D56         S
22       0            330923      8.0292          NaN         Q
23       0            113788     35.5000           A6         S
24       1            349909     21.0750          NaN         S
25       5            347077     31.3875          NaN         S
26       0              2631      7.2250          NaN         C
27       2             19950    263.0000  C23 C25 C27         S
28       0            330959      7.8792          NaN         Q
29       0            349216      7.8958          NaN         S
..     ...               ...         ...          ...       ...
861      0             28134     11.5000          NaN         S
862      0             17466     25.9292          D17         S
863      2          CA. 2343     69.5500          NaN         S
864      0            233866     13.0000          NaN         S
865      0            236852     13.0000          NaN         S
866      0      SC/PARIS 2149    13.8583          NaN         C
867      0          PC 17590     50.4958          A24         S
868      0            345777      9.5000          NaN         S
869      1            347742     11.1333          NaN         S
870      0            349248      7.8958          NaN         S
871      1             11751     52.5542          D35         S
872      0               695      5.0000  B51 B53 B55         S
873      0            345765      9.0000          NaN         S
```

```
874    0      P/PP 3381   24.0000       NaN    C
875    0            2667    7.2250       NaN    C
876    0            7534    9.8458       NaN    S
877    0          349212    7.8958       NaN    S
878    0          349217    7.8958       NaN    S
879    1           11767   83.1583       C50    C
880    1          230433   26.0000       NaN    S
881    0          349257    7.8958       NaN    S
882    0            7552   10.5167       NaN    S
883    0  C.A./SOTON 34068   10.5000       NaN    S
884    0   SOTON/OQ 392076    7.0500       NaN    S
885    5          382652   29.1250       NaN    Q
886    0          211536   13.0000       NaN    S
887    0          112053   30.0000       B42    S
888    2       W./C. 6607   23.4500       NaN    S
889    0          111369   30.0000      C148    C
890    0          370376    7.7500       NaN    Q

[891 rows x 12 columns]
```

In [ ]:

8