titanic-1

October 17, 2018

```
In [104]: %matplotlib inline
```

```
In [105]: from fastai.imports import *
          from fastai.structured import *

          from pandas_summary import DataFrameSummary
          from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
```

```
In [106]: PATH = "data/titanic/"
```

```
In [78]: !ls {PATH}
```

```
gender_submission.csv  test.csv  titanic.zip  train.csv
```

```
In [79]: df_raw = pd.read_csv(f'{PATH}train.csv', low_memory=False)
```

```
In [80]: df_raw.shape
```

```
Out[80]: (891, 12)
```

```
In [81]: def cats_train(df):
             for n, c in df.items():
                 if is_string_dtype(c): df[n] = c.astype('category').cat.as_ordered()
```

```
In [82]: cats_train(df_raw)
```

```
In [83]: df_raw.isnull().sum().sort_index()/len(df_raw)
```

```
Out[83]: Age            0.198653
         Cabin          0.771044
         Embarked       0.002245
         Fare           0.000000
         Name           0.000000
         Parch          0.000000
         PassengerId    0.000000
         Pclass         0.000000
         Sex            0.000000
         SibSp          0.000000
         Survived       0.000000
         Ticket         0.000000
         dtype: float64
```

```
In [84]: age = df_raw['Age'].dropna()

In [85]: df_raw['Age'].fillna(np.median(age), inplace=True)

In [86]: def cat_codes(df):
             for n, c in df.items():
                 if not is_numeric_dtype(df[n]):
                     df[n] = df[n].cat.codes+1

In [ ]:

In [88]: def train_valid(df):
             msk = np.random.rand(len(df)) < 0.90
             res = [df[msk], df[~msk]]
             return res

In [89]: df_raw_train, df_raw_valid = train_valid(df_raw)

In [ ]: cat_codes(df_raw_train)
        cat_codes(df_raw_valid)

In [90]: print(df_raw_train.shape,
         df_raw_valid.shape)

(792, 12) (99, 12)


In [91]: df_raw_train_y = df_raw_train['Survived']
         df_raw_train.drop('Survived', axis=1, inplace=True)

/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/pandas/core/frame.py:3697: Setti
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  errors=errors)


In [92]: df_raw_valid_y = df_raw_valid['Survived']
         df_raw_valid.drop('Survived', axis=1, inplace=True)

In [96]: df_raw_train.head()

Out[96]:    PassengerId  Pclass  Name  Sex   Age  SibSp  Parch  Ticket    Fare  Cabin  \
        0            1       3   109    2  22.0      1      0     524  7.2500      0
        2            3       3   354    1  26.0      0      0     670  7.9250      0
        3            4       1   273    1  35.0      1      0      50 53.1000     56
        5            6       3   555    2  28.0      0      0     276  8.4583      0
        6            7       1   516    2  54.0      0      0      86 51.8625    130

           Embarked
```

```
        0          3
        2          3
        3          3
        5          2
        6          3
```

In [97]: m = RandomForestRegressor(n_jobs=-1, n_estimators=30, max_features='sqrt', oob_score=
         m.fit(df_raw_train, df_raw_train_y)

Out[97]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
                   max_features='sqrt', max_leaf_nodes=None,
                   min_impurity_decrease=0.0, min_impurity_split=None,
                   min_samples_leaf=1, min_samples_split=2,
                   min_weight_fraction_leaf=0.0, n_estimators=30, n_jobs=-1,
                   oob_score=True, random_state=None, verbose=0, warm_start=False)

In [98]: m.score(df_raw_train, df_raw_train_y)

Out[98]: 0.9170690198127935

In [99]: m.score(df_raw_valid, df_raw_valid_y)

Out[99]: 0.4258080808080809

In [101]: m.oob_score_

Out[101]: 0.4293453058016202

In [108]: %%bash
          jupyter nbconvert --to=latex 'Index.ipynb' --post=pdf

Traceback (most recent call last):
  File "/home/diablo/anaconda3/envs/fastai/bin/jupyter-nbconvert", line 10, in <module>
    sys.exit(main())
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/jupyter_core/application
    return super(JupyterApp, cls).launch_instance(argv=argv, **kwargs)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/applica
    app.initialize(argv)
  File "<decorator-gen-7>", line 2, in initialize
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/applica
    return method(app, *args, **kwargs)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/nbconvert/nbconvertapp.p
    super(NbConvertApp, self).initialize(argv)
  File "<decorator-gen-6>", line 2, in initialize
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/applica
    return method(app, *args, **kwargs)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/jupyter_core/application
    self.parse_command_line(argv)
  File "<decorator-gen-4>", line 2, in parse_command_line
```

```
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/applica
    return method(app, *args, **kwargs)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/applica
    self.update_config(self.cli_config)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/configu
    self._load_config(config)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/config/configu
    warn(msg)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/contextlib.py", line 88, in __exit__
    next(self.gen)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/traitlets.py"
    self.notify_change(change)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/traitlets.py"
    c(change)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/nbconvert/nbconvertapp.p
    self.postprocessor_factory = import_item(new)
  File "/home/diablo/anaconda3/envs/fastai/lib/python3.6/site-packages/traitlets/utils/importst
    return __import__(parts[0])
ModuleNotFoundError: No module named 'pdf'



        ---------------------------------------------------------------------------

        CalledProcessError                        Traceback (most recent call last)

        <ipython-input-108-3cd60aa8115d> in <module>
    ----> 1 get_ipython().run_cell_magic('bash', '', "jupyter nbconvert  --to=latex 'Index.ipyn


        ~/anaconda3/envs/fastai/lib/python3.6/site-packages/IPython/core/interactiveshell.py in
       2319             magic_arg_s = self.var_expand(line, stack_depth)
       2320             with self.builtin_trap:
    -> 2321                 result = fn(magic_arg_s, cell)
       2322             return result
       2323


        ~/anaconda3/envs/fastai/lib/python3.6/site-packages/IPython/core/magics/script.py in na
        140             else:
        141                 line = script
    --> 142             return self.shebang(line, cell)
        143
        144         # write a basic docstring:


        <decorator-gen-109> in shebang(self, line, cell)
```

```
~/anaconda3/envs/fastai/lib/python3.6/site-packages/IPython/core/magic.py in <lambda>(
185      # but it's overkill for just that one bit of state.
186      def magic_deco(arg):
--> 187          call = lambda f, *a, **k: f(*a, **k)
188
189          if callable(arg):


~/anaconda3/envs/fastai/lib/python3.6/site-packages/IPython/core/magics/script.py in s
243              sys.stderr.flush()
244          if args.raise_error and p.returncode!=0:
--> 245              raise CalledProcessError(p.returncode, cell, output=out, stderr=err)
246
247      def _run_script(self, p, cell, to_close):


CalledProcessError: Command 'b"jupyter nbconvert  --to=latex 'Index.ipynb' --post=pdf\n
```

In [ ]: