HOMEWORK_Feb19_Elda

February 19, 2020

```python
[1]: import numpy as np
     import pandas as pd
     from sklearn.model_selection import train_test_split
     from sklearn import metrics
     from sklearn.impute import SimpleImputer
     from sklearn.ensemble import RandomForestClassifier
```

Using TensorFlow backend.

```python
[51]: from sklearn.impute import SimpleImputer

      # Imputation
      my_imputer = SimpleImputer()
      imputed_X_train= pd.DataFrame(my_imputer.fit_transform(train))
      imputed_X_test= pd.DataFrame(my_imputer.transform(test))

      # Imputation removed column names; put them back
      imputed_X_train.columns = train.columns
      imputed_X_train.index = train.index
      imputed_X_test.columns = test.columns
      imputed_X_test.index = test.index
```

```python
[52]: # Fit Model
      from sklearn.ensemble import RandomForestClassifier

      model= RandomForestClassifier(random_state=1)
      model.fit(imputed_X_train, target)
```

```
[52]: RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                             criterion='gini', max_depth=None, max_features='auto',
                             max_leaf_nodes=None, max_samples=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=100,
                             n_jobs=None, oob_score=False, random_state=1, verbose=0,
                             warm_start=False)
```

```python
[41]: model.classes_
```

```
[41]: array([0, 1])
```

```
[53]: preds_test_2 = model.predict_proba(imputed_X_test)
```

```
[54]: preds_test_2
```

```
[54]: array([[0.93, 0.07],
             [0.83, 0.17],
             [0.9 , 0.1 ],
             ...,
             [0.76, 0.24],
             [0.98, 0.02],
             [0.83, 0.17]])
```

```
[55]: preds_test_2.T[1]
```

```
[55]: array([0.07, 0.17, 0.1 , ..., 0.24, 0.02, 0.17])
```

```
[59]: test['encounter_id']
```

```
[59]: 0            2
      1            5
      2            7
      3            8
      4           10
              ...
      39303    131035
      39304    131037
      39305    131039
      39306    131041
      39307    131050
      Name: encounter_id, Length: 39308, dtype: int64
```

```
[61]: output = pd.DataFrame({'encounter_id': test['encounter_id'],
                             'hospital_death': preds_test_2.T[1]})

      output.to_csv('trial_bad', index=False)
```

```
[ ]:
```