PCA visualizations

January 29, 2016

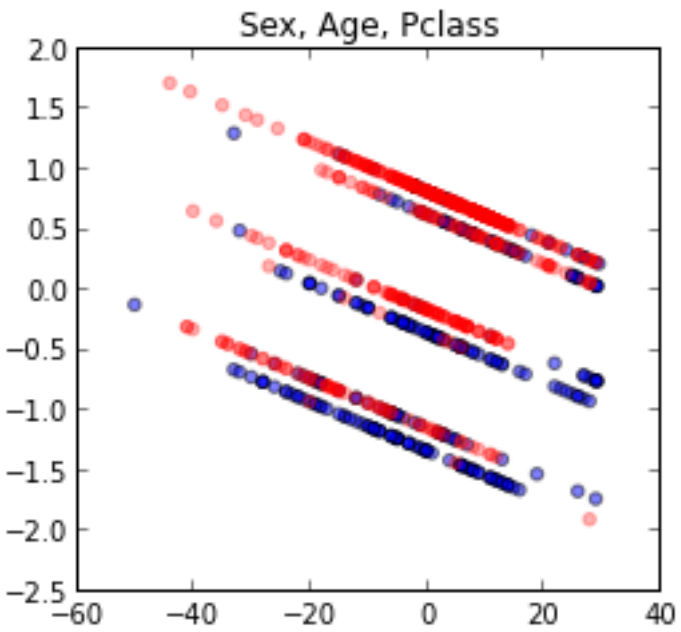
0.1 Titanic Machine Learning from Disaster- Kaggle competition

http://www.kaggle.com/c/titanic-gettingStarted

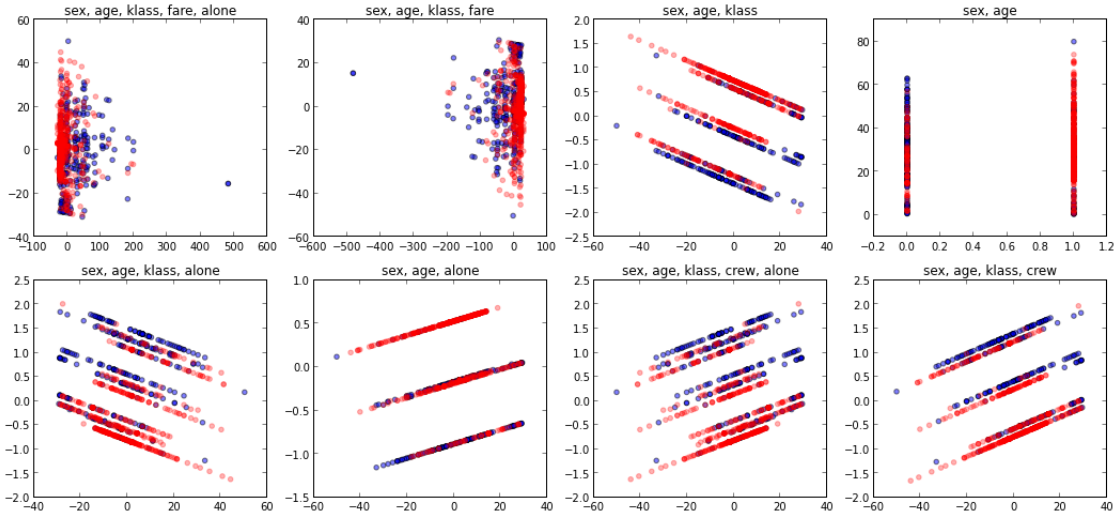
In [6]:
```
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA

train_data = pd.read_csv('train.csv')
test_data = pd.read_csv('test.csv')

train_data['type'] = 'train'
test_data['type'] = 'test'

test_data['Survived'] = np.nan

all_data = train_data.append(test_data, ignore_index=True)
```

In [7]:
```
from predict import mungling

data = mungling(all_data)
```

In [8]:
```
plt.figure(figsize(18, 8))

def show_survival(data, fields):
    if len(fields) > 2:
        pca = PCA(n_components=2)
        decomp = pca.fit_transform(data[data.type == 'train'][fields])
    else:
        decomp = np.array(data[fields])

    surv = decomp[(data.type == 'train') & (data.survived == 1)]
    nosurv = decomp[(data.type == 'train') & (data.survived == 0)]

    plt.scatter(surv[:,0:1], surv[:,1:2], alpha=0.5)
    plt.scatter(nosurv[:,0:1], nosurv[:,1:2], alpha=0.3, color='r')
    plt.title(', '.join(fields))

train_data.Sex.replace({'male': 1, 'female': 0}, inplace=True)
train_data['survived'] = train_data.Survived


plt.subplot(2, 4, 1)
show_survival(train_data[train_data.Age.notnull()], ['Sex', 'Age', 'Pclass'])
```




```
In [9]: data.sex.replace({'male': 1, 'female': 0}, inplace=True)

        plt.subplot(2, 4, 1)
        show_survival(data, ['sex', 'age', 'klass', 'fare', 'alone'])
        plt.subplot(2, 4, 2)
        show_survival(data, ['sex', 'age', 'klass', 'fare'])
        plt.subplot(2, 4, 3)
        show_survival(data, ['sex', 'age', 'klass'])
        plt.subplot(2, 4, 4)
        show_survival(data, ['sex', 'age'])
        plt.subplot(2, 4, 5)
        show_survival(data, ['sex', 'age', 'klass', 'alone'])
        plt.subplot(2, 4, 6)
        show_survival(data, ['sex', 'age', 'alone'])
        plt.subplot(2, 4, 7)
        show_survival(data, ['sex', 'age', 'klass', 'crew', 'alone'])
        plt.subplot(2, 4, 8)
        show_survival(data, ['sex', 'age', 'klass', 'crew'])
```


sex, age, klass, fare, alone
sex, age, klass, fare
sex, age, klass
sex, age
sex, age, klass, alone
sex, age, alone
sex, age, klass, crew, alone
sex, age, klass, crew