

Now we will focus on Senator Voting data. This data provides information about senator vote x and senator political affiliation y . We provide you with four different vectors (a_1, a_2, a_3, a_4) precomputed by the EECS127 staff. Each of these vectors can be used to define a linear function $f_a : x \rightarrow a^T x$.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

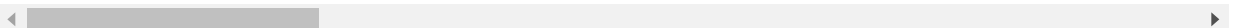
```
In [2]: senator_df = pd.read_csv('data_matrix.csv', index_col=0)
a_vectors = np.loadtxt('vectors.txt').reshape(4, 542)
affiliation_file = open("politician_labels.txt", "r")
affiliations = np.array([line.split("\n")[0].split(" ")[1] for line in affiliation_file.readlines()])
```

```
In [3]: senator_df.head()
```

Out[3]:

	bill_type	bill_name	bill_ID	missing_votes	Arlen Specter (PA)	Barack H. Obama (IL)	BarbaraA Mikulski(MD)	BarbaraBoxer(CA)
2	Appropriations_Transit Security Amendment_3866			0.0	1.0	1.0	1.0	
3	Budget_Spending_and_Taxes_2007 Budget Resoluti...			0.0	1.0	-1.0	-1.0	-
4	"Budget, Spending and Taxes_Debt Limit Increas...			0.0	1.0	-1.0	-1.0	-
5	"Budget, Spending and Taxes_Education Funding ...			0.0	-1.0	1.0	1.0	
6	Budget, Spending and Taxes_Reinstate Pay-As-Yo...			0.0	-1.0	1.0	1.0	

5 rows × 102 columns



Now that we have collected all the data, we will clean the senator voting data and convert it to numerical format

```
In [49]: X = np.array(senator_df.values[:-1, 2:], dtype='float64')
```

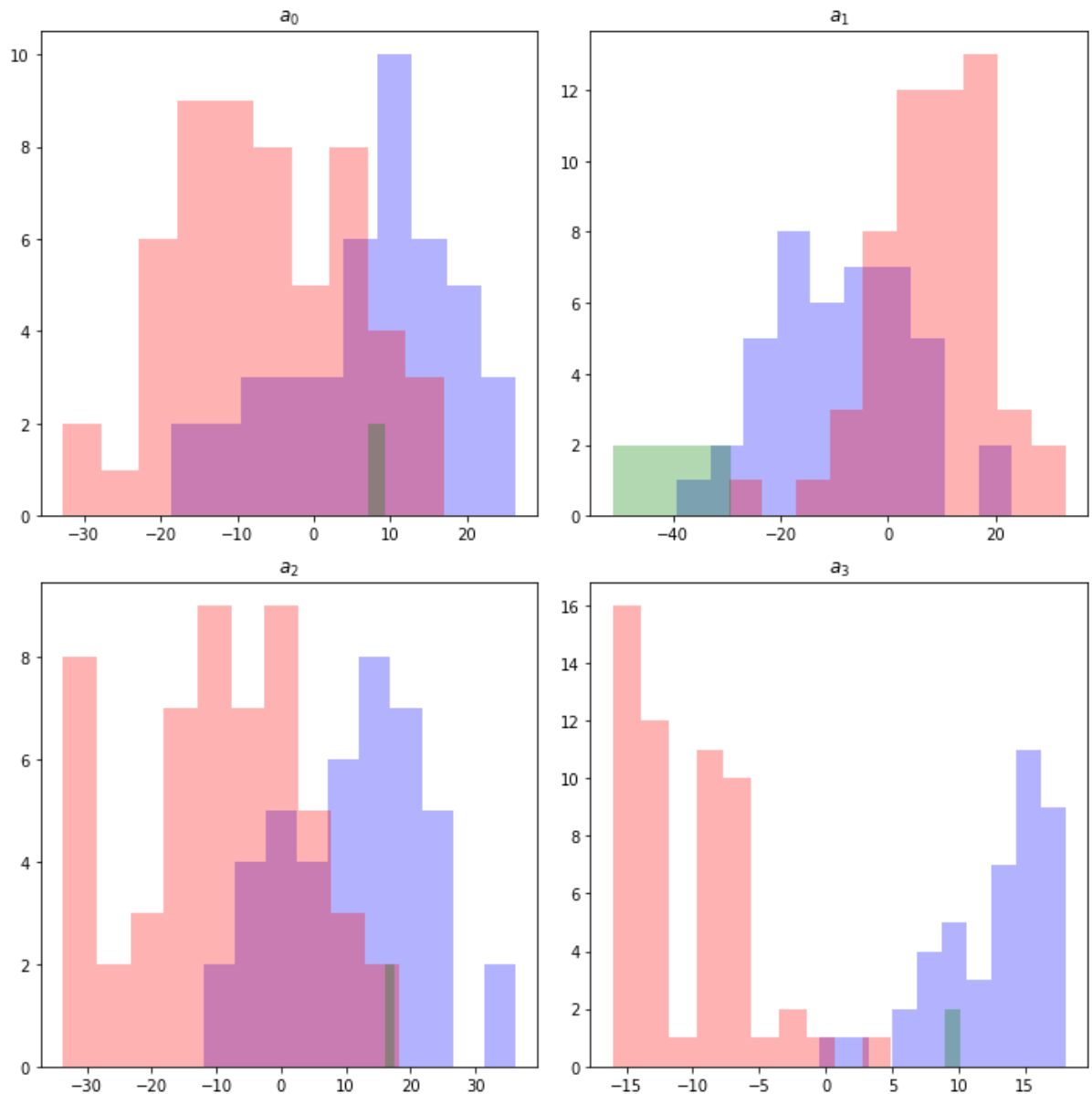
```
In [50]: # TODO: Center the data matrix X by removing to each column its mean
X_bar = X.mean(axis=1)
for i in range(X.shape[1]):
    X[:, i] = X[:, i] - X_bar
len(X[0])
```

Out[50]: 100

```
In [51]: # TODO: compute for each vector a the score of each senator
senator_scores = np.zeros(shape=(4, 100))
print(len(senator_scores), len(senator_scores[0]))
for i in range(4):
    senator_scores[i] = a_vectors[i].dot(X)
```

4 100

```
In [52]: # Then we help you visualizing the scores with the library matplotlib
f, axarr = plt.subplots(2, 2, figsize=(10, 10))
for i in range(4):
    axarr[i // 2, i % 2].hist(senator_scores[i], affiliations == "Blue", color="Blue", alpha=0.3)
    axarr[i // 2, i % 2].hist(senator_scores[i], affiliations == "Red", color="Red", alpha=0.3)
    axarr[i // 2, i % 2].hist(senator_scores[i], affiliations == "Yellow", color="Green", bins=1, alpha=0)
    axarr[i // 2, i % 2].set_title(r'$a_{\text{' + str(i) + '}}$')
plt.tight_layout()
plt.show()
```



In []:

