mysterious_data_sol

November 19, 2021

[ ]:

```
%matplotlib inline
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
```

[ ]:

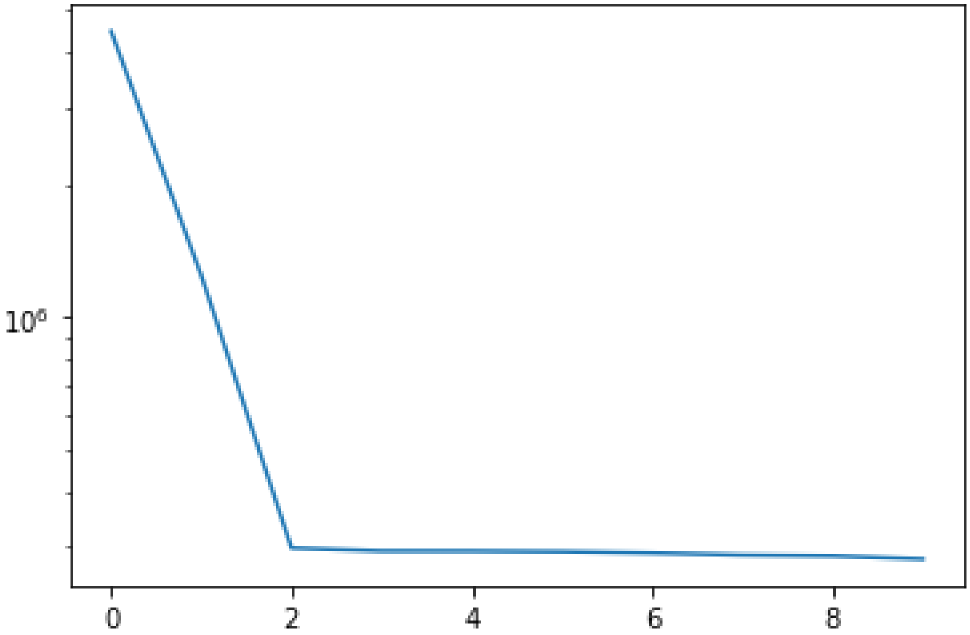
```
# Load the data matrix
A = np.loadtxt('mysterious_data__.txt')
n,d = A.shape
print(f'The matrix A contains {n} points in dimension {d}')
```

```
The matrix A contains 3000 points in dimension 1000
```

Each row of A corresponds to a datapoint.

[ ]:

```
Ac = A - A.mean(0)
AAT = Ac.T @ Ac
sp, vec = np.linalg.eigh(AAT,)
plt.plot(sp[::-1][:10])
plt.yscale('log')
```

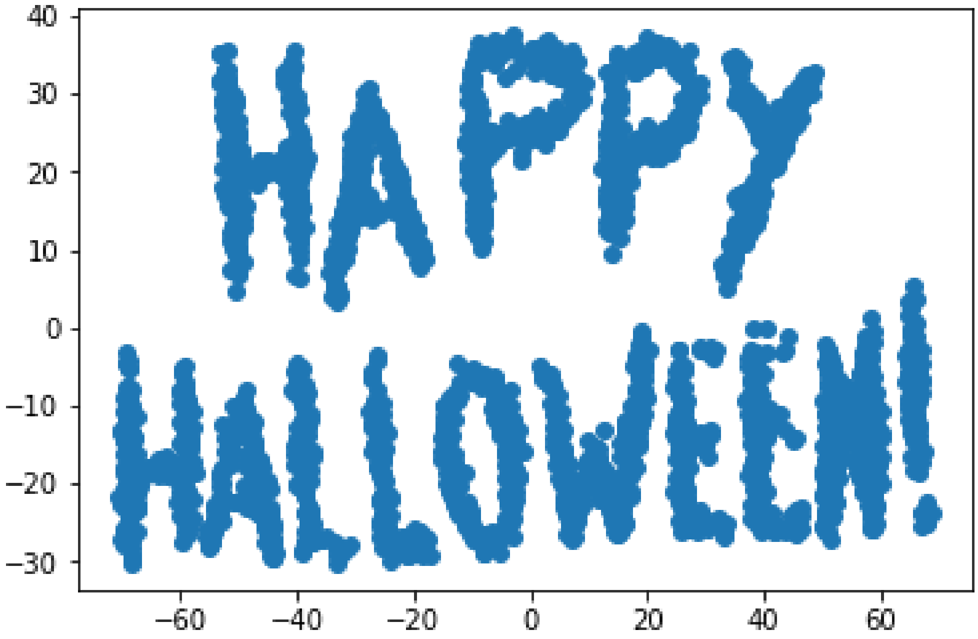


```
[ ]: f = Ac @ vec[:, -2:]
     f.shape
```

```
[ ]: (3000, 2)
```

```
[ ]: plt.scatter(f[:, 1],  - f[:, 0])
```

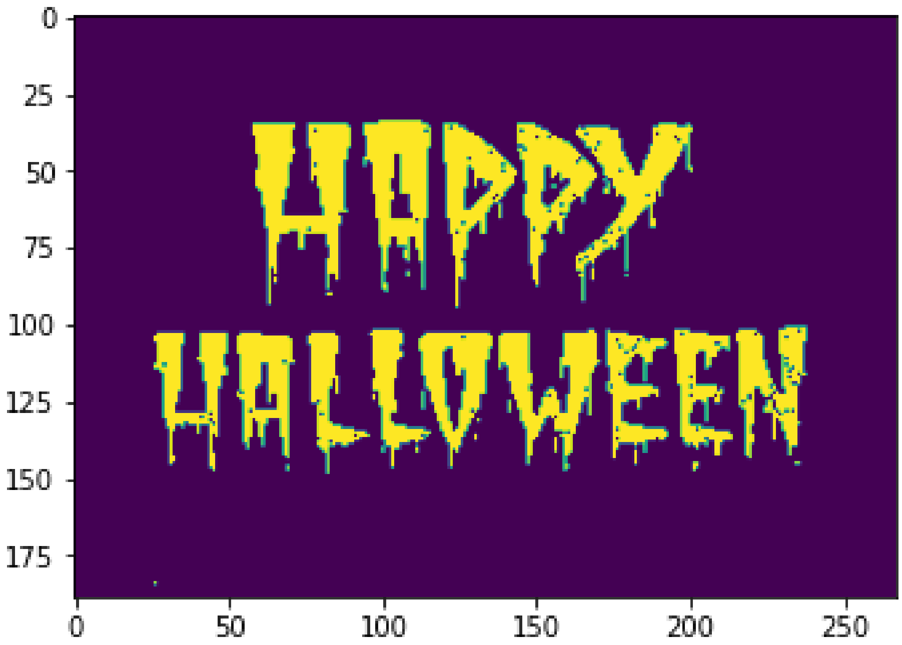
```
[ ]: <matplotlib.collections.PathCollection at 0x7f9ef6a34c90>
```

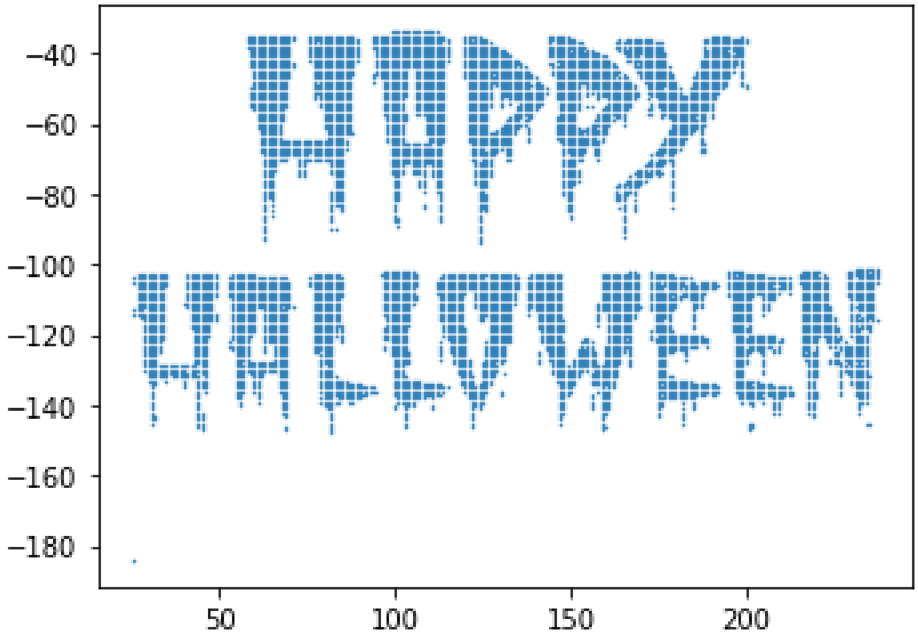


1 How was the data created?

```
[ ]: from matplotlib import image
     import matplotlib.pyplot as plt
     import numpy as np
     # load image as pixel array

     image = image.imread('images.jpg').sum(-1)
     image = 1 - image / image.max()
     image_ = image > 0.85
     plt.imshow(image_)

     ## get points
```

```
xs, ys = np.where(image > 0.85)
plt.figure()
ax = plt.subplot(111)
ax.scatter(ys, -xs, s=0.2)
ax.set_aspect('equal')
n_points = xs.shape[0]

## hide point in higher dimenion - with smaller random noise
dim = 1000
image_hd = 0.01 * np.random.randn(n_points, dim)
image_hd[:, 0] += 1e4 * xs
image_hd[:, 1] += 1e4 * ys

data = image_hd

## change basis using an orthogonal matrix
from scipy.stats import ortho_group  # Requires version 0.18 of scipy

P = ortho_group.rvs(dim=dim)
data = image_hd @ P
```

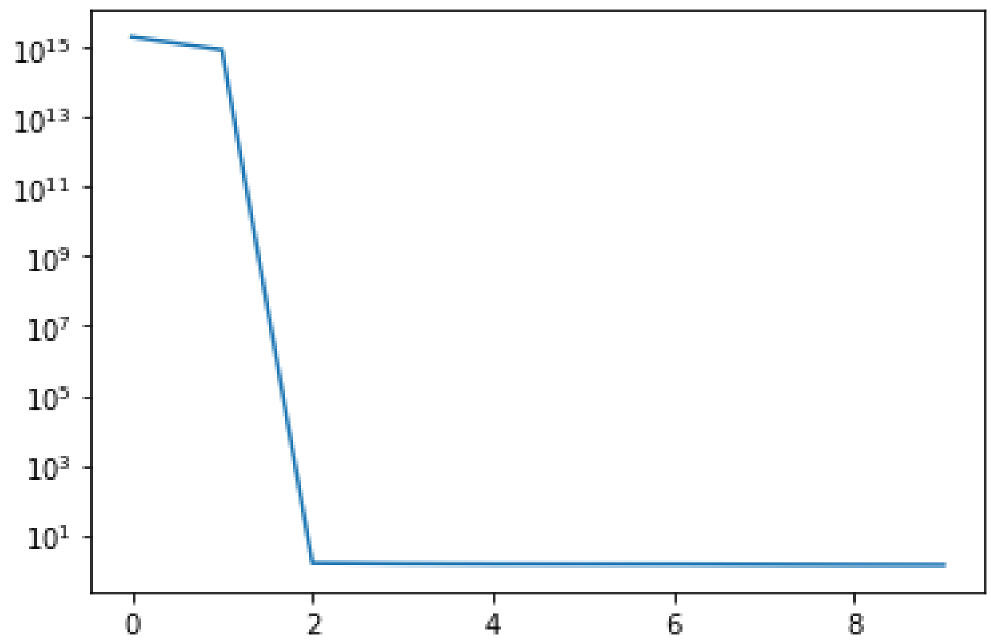




[ ]:

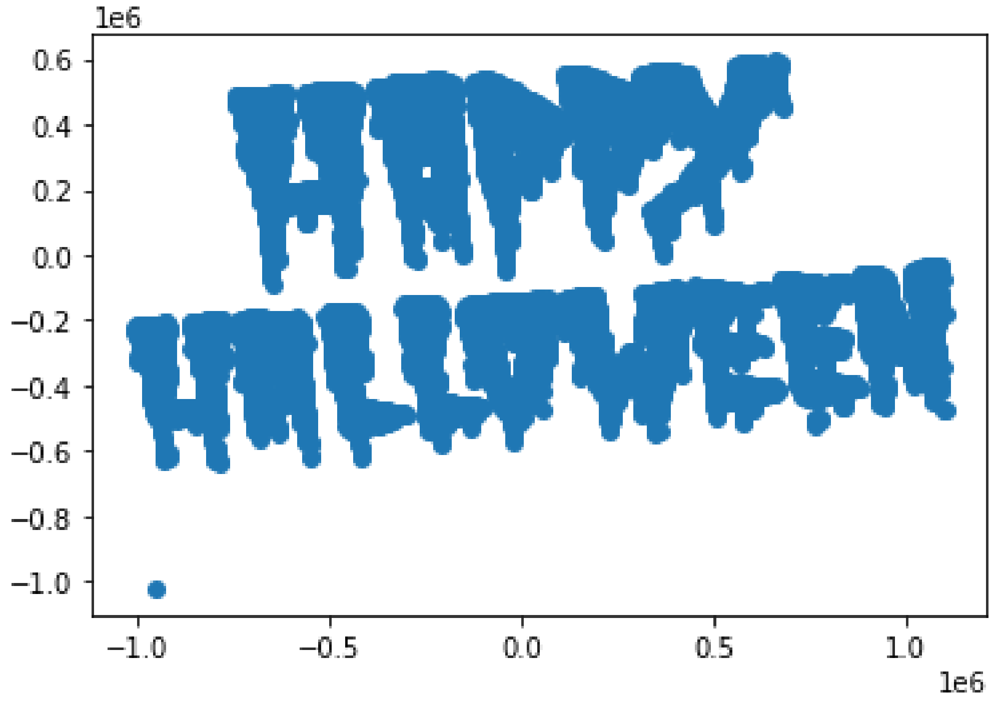
```
A = data
n,d = A.shape
print(f'The matrix A contains {n} points in dimension {d}')
Ac = A - A.mean(0)
AAT = Ac.T @ Ac
sp, vec = np.linalg.eigh(AAT,)
plt.plot(sp[::-1][:10])
plt.yscale('log')
```

```
The matrix A contains 7391 points in dimension 1000
```



```
[ ]: f = Ac @ vec[:, -2:]
     plt.scatter(- f[:, 1],  f[:, 0])
```

```
[ ]: <matplotlib.collections.PathCollection at 0x7fe721264d10>
```



$$6$$

[ ]:

[ ]: