

# AM 170B HW 2

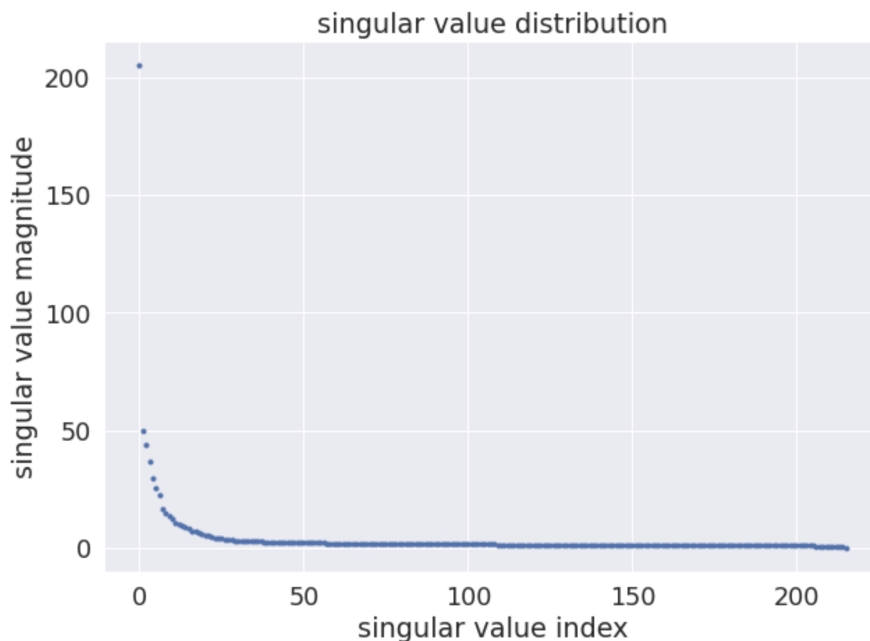
Henry Charlton

2022 April 17

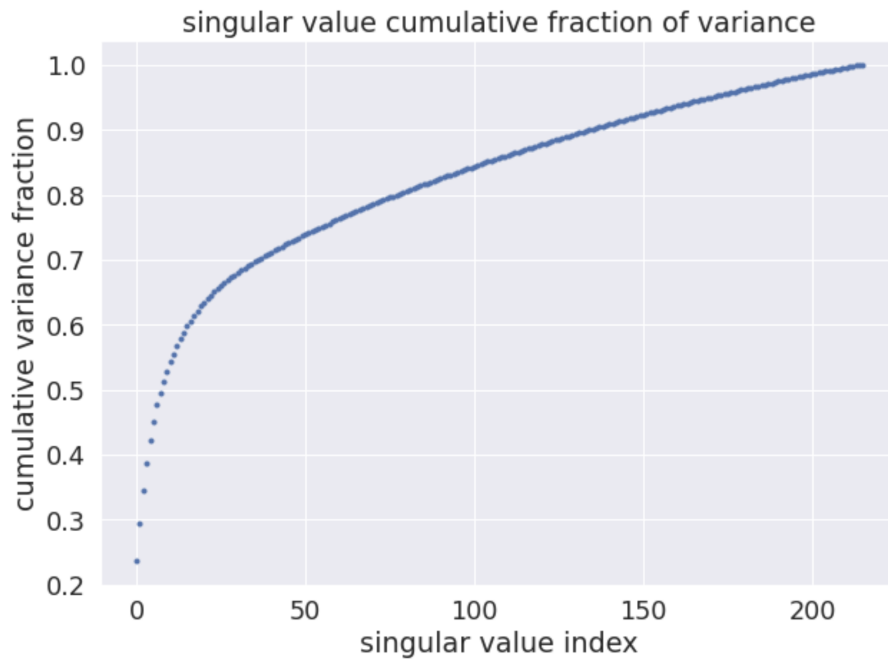
## 1 Principle Component Analysis on Ovarian Cancer Gene Expression

For this analysis we use the Numpy SVD function to implement principal component analysis on mean centered data regarding the gene expressions of 216 patients with respect to whether they have ovarian cancer or not.

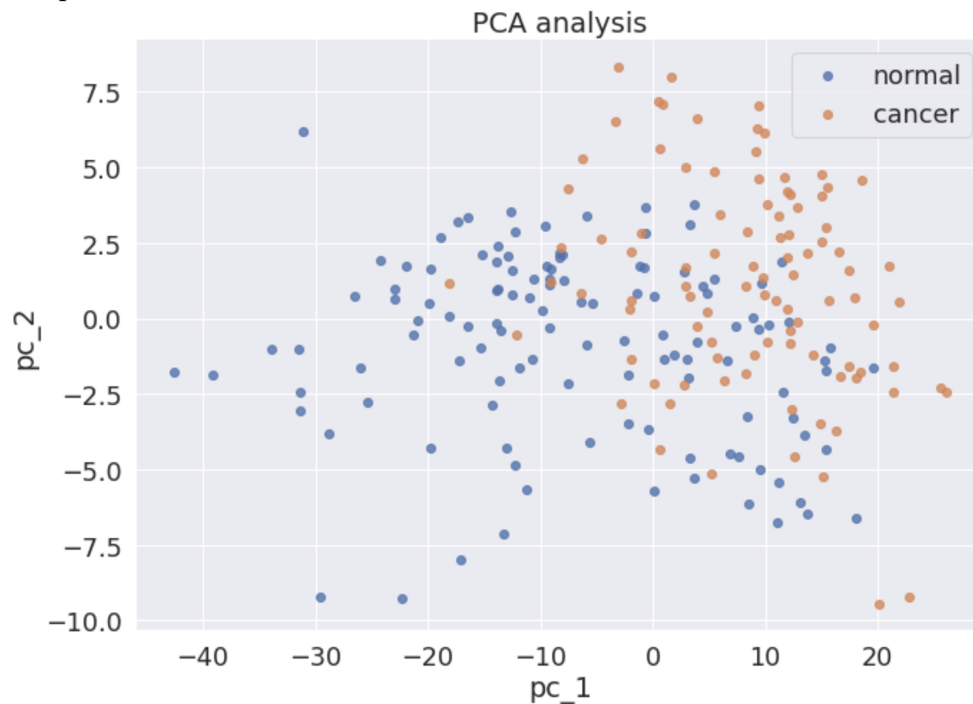
The plot below shows the distribution of singular values, highlighting that the first 10 or so are vastly greater in magnitude than the rest and thus contain the dominant modes.



Furthermore we can see in the plot below that after around the first 20 singular values, much less information is gained by adding the next one than before, ie there appears to be an elbow around the 20th singular value.

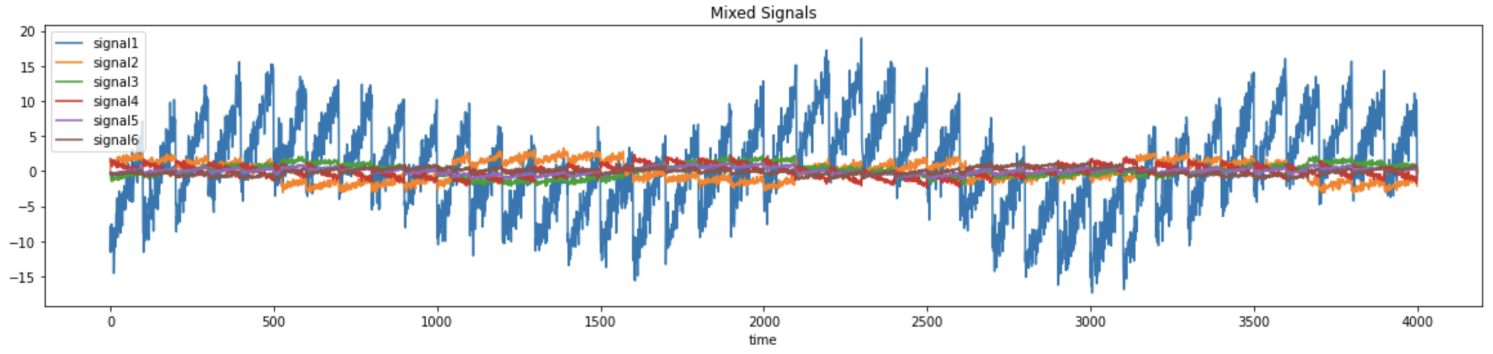


Finally, using our defined PCA function, we can see the separation and projection on the first two principal components of the data. Note that there is decent separation between cancer and normal samples.



## 2 Independent Component Analysis on Mixed Signals

For this analysis we use the scikit-learn PCA and ICA functions to decompose a set of 6 mixed signals and compare the results. Below is the plot of the mixed signals.



The below plot shows the ICA blind transformed source signals, showing a good separation of the mixed signal into four source signals that are easily described by the following mathematical functions:

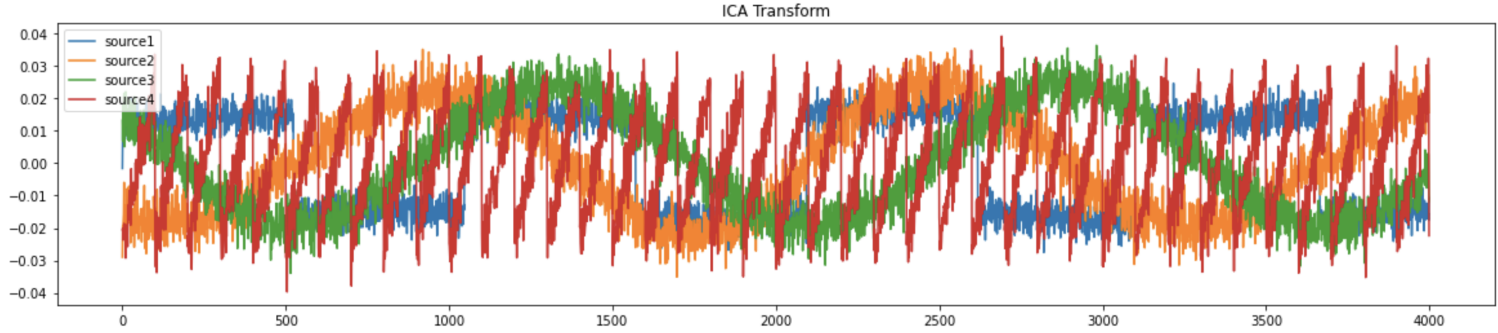
$$f_1(t) \approx 0.0314 \sin(0.0043t)$$

$$f_2(t) \approx 0.0314 \sin(0.0043t - \frac{\pi}{2})$$

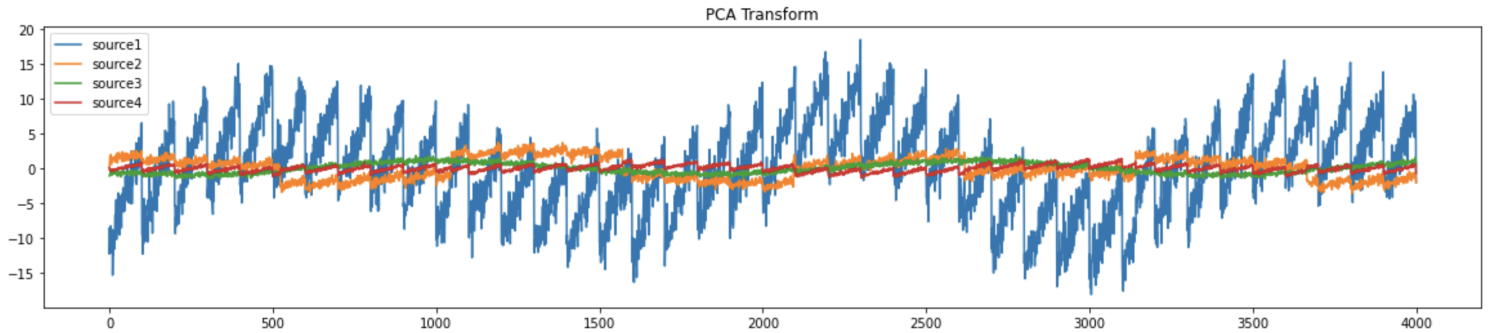
$$f_3(t) \approx \sum_{n=1}^{40} \frac{0.0314 \sin(-0.00628(2n+1)t)}{2n+1}$$

$$f_4(t) \approx \sum_{n=1}^{40} \frac{0.0314 \sin(-0.00314(2n)t)}{2n}$$

Note that we found that four source signals had the best approximation into simple functions.



Below a plot of the PCA transformed source signals is shown. This is obviously an inferior set of source signals prima facie, owing to the fact that there is a sine wave mixed with a sawtooth function, whereas in the ICA source signals those were successfully separated into two different functions.



### 3 Course Project

I will be collaborating with Prof. Vanessa Jonsson to design a genomics focused course project. Correspondence is en route via email for this to take place.