Ross Spencer

Question 1

1. Will PCA be effective and keeping the two clusters separated in the reduced dimensionality data? Why or why not? If yes, state what characteristics of the data set allow PCA to be effective. If no, state what characteristics of the data set cause PCA to fail.

For the first dataset, PCA will be effective. The most variance seems to be on the y-axis. Since the data is elliptical in both clusters in the direction of the PC1 vector (the y-axis) with white space between them, when projected down to 1 dimension along PC1 the clusters will remain separated.

For the second dataset, PCA will not be effective. The most variance is along the y-axis, and when projected onto the y-axis (our PC1 vector), the clusters are smushed onto one another, making them inseparable. The most useful feature is along the x-axis, which is only the 2nd most variance and will thus be neglected by PCA.

For the third dataset, PCA will not be effective. The inner cluster of data is mostly inside of the other, so any line that you project it onto will make the clusters inseparable.

2. Will LDA be effective and keeping the two clusters separated in the reduced dimensionality data? Why or why not? If yes, state what characteristics of the data set allow LDA to be effective. If no, state what characteristics of the data set cause LDA to fail.

For the first dataset, LDA will be effective. The means between the clusters are very separate, and the classes themselves are very separable.

For the second dataset, LDA will be effective. The means between the clusters are very separate, and LDA will separate them by projecting down onto the x-axis as desired instead of projecting onto y, avoiding what would cause PCA to fail on this dataset.

For the third dataset, LDA will not be effective as the 2 clusters have overlapping means. This makes it hard for the algorithm to maximize the distance between their means while minimizing the spread.

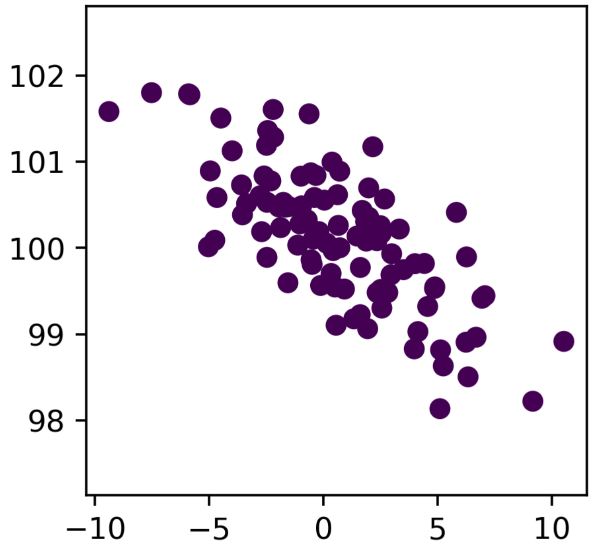
Question 2

1. To obtain our whitened version of x, we want to:

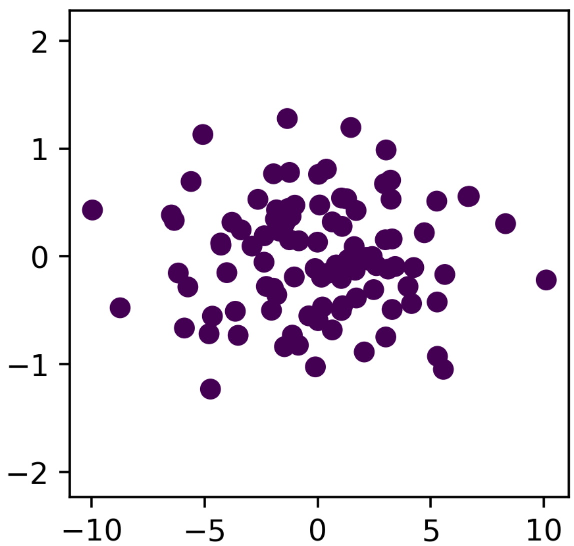
subtract the mean from X, find the correlation matrix of X, compute the eigenvectors & eigenvalues of the correlation matrix, sort the eigenvectors from largest eigenvalue to smallest, pick dimensionality, and set A equal to the eigenvectors with the largest eigenvalues.

So, our formula for z is: $\mathbf{z} = (\mathbf{\Lambda}^{-1/2}\mathbf{U}^{\top}(\mathbf{X}^{(i)}-\boldsymbol{\mu}))$, where i is the index of the x we're trying to whiten ($x = X^{(i)}$), lambda is the diagonal matrix of eigenvalues, U is an orthogonal rotation matrix composed of eigenvectors of the correlation matrix, and mu is the mean of our dataset X. Source: http://ace.cs.ohiou.edu/~razvan/courses/dl6890/lecture03.pdf

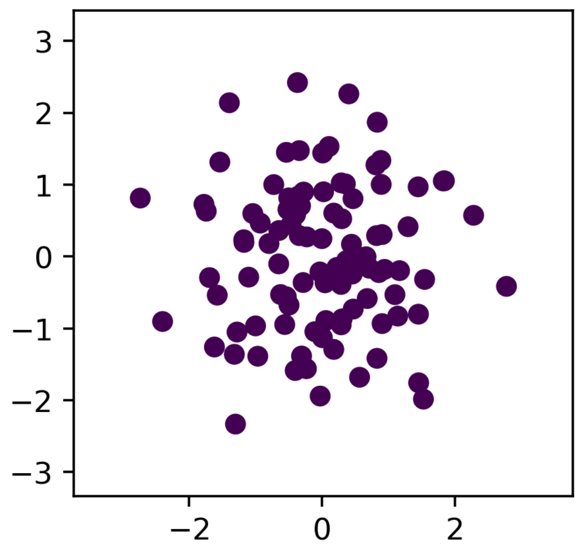
2.



Pictured above is the original dataset with mean $[1,100]^T$ and covariance matrix [[13,-2], [-2,.001]].



Our dataset after decorrelation but before whitening.



Our dataset after both decorrelation and whitening.

The first graph shows the original dataset. Our covariance after PCA without whitening is: array([[1.08151390e+01, 6.85168768e-18], [6.85168768e-18, 2.73218478e-01]]). After whitening, the covariance matrix becomes array([[1.00000000e+00, -2.87675398e-17], [-2.87675398e-17, 1.00000000e+00]]), where each attribute has an equal variance of 1 and the covariance is effectively 0. You would want to apply data whitening to data with features of widely varying range and variance, as it makes the data a lot easier to work with since it becomes more elliptical and gets rid of noise, reducing redundancy in the data.