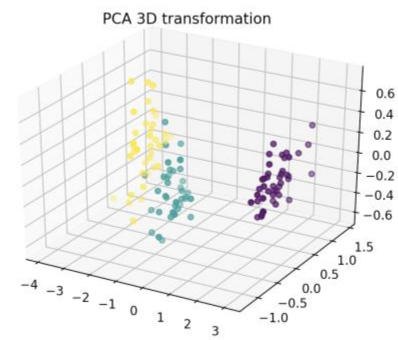
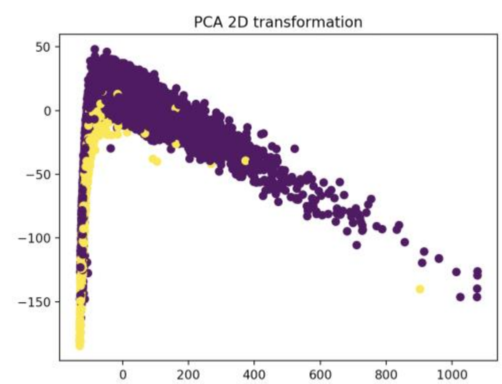
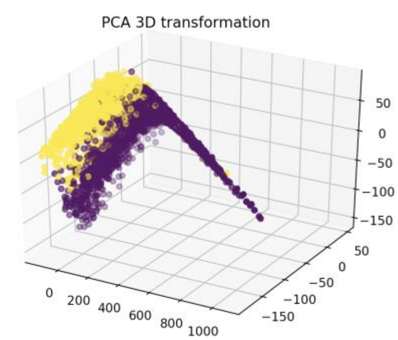
Zachary Panzarino
903305160
HW3

1. *PCA*
   a. **Iris**

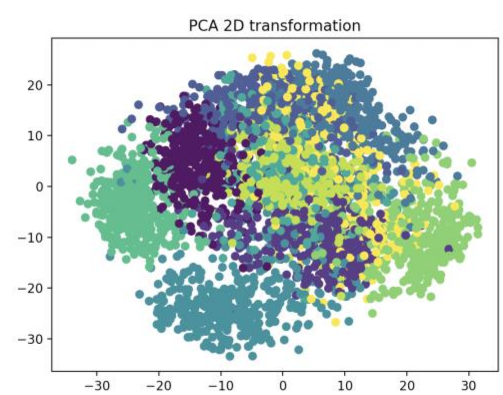


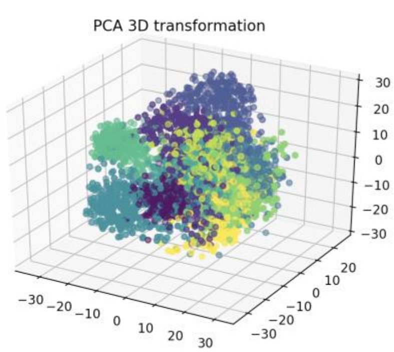
HTRU2





Digits





   b. **Iris**
   For both 2D and 3D, using PCA for data transformation would improve classification. Each of the 3 classes are fairly distinct, indicating that the PCA was

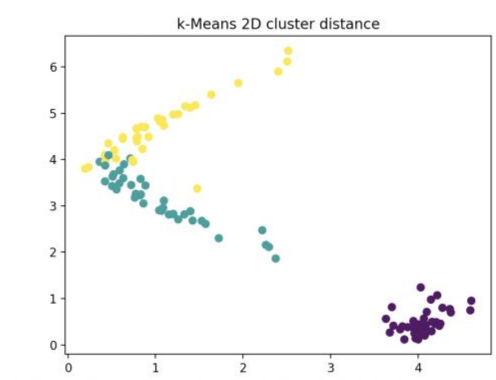
able to remove some noise from the data. For this dataset, both the 2D and 3D transformations seem to effectively distinguish the purple label from the others. In addition, the 3D performs even better than the 2D because it seems to add some additional separation between the yellow and green labels.

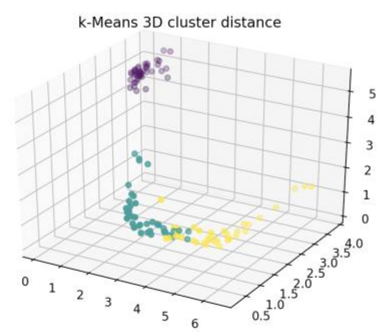
HTRU2
In the 2D case, PCA would most likely reduce classification performance. While some of the purple dots are clearly different than the yellow, the section with the majority of yellow dots also has many purple dots in the same vicinity. This indicates that another dimension has key information about this dataset. The 3D case, however, would definitely increase classification performance. In 3D, the labels are clearly distinguished, indicating that the PCA is effectively removing some noise in this case.

Digits
Both the 2D and 3D cases would reduce classification performance. Both cases show many labels that are not very distinguished from each other. In the 3D model, some labels, such as the green and teal in the bottom left, begin to distinguish themselves from the rest. PCA may potentially be an effective tool for classification in higher dimensions, as the data is very noisy in these dimensions.

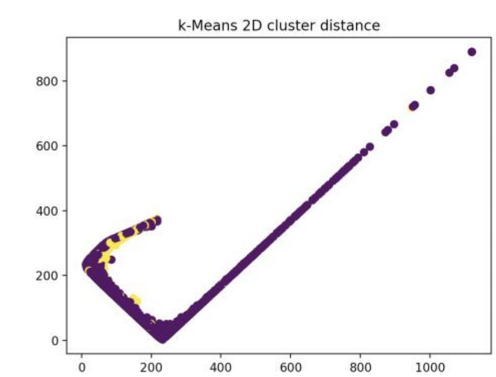
2. *k-Means*

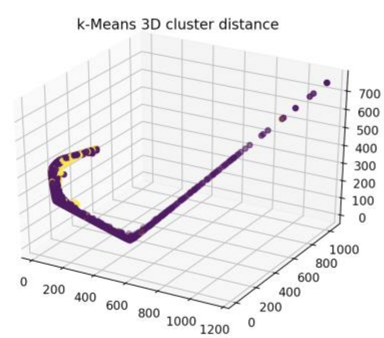
   a. **Iris**



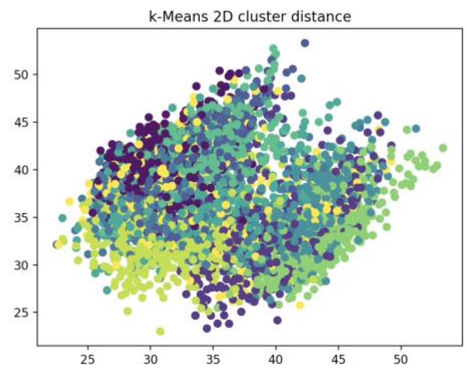


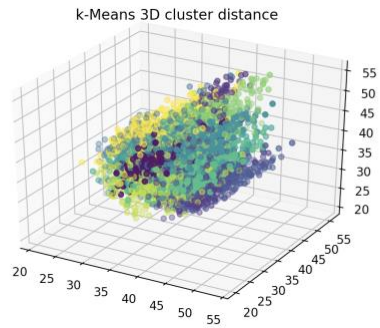
HTRU2





Digits





b. Iris

k-Means seems to be better for transforming the data in both of the cases. While PCA does good in both cases, k-Means seems to do slightly better when it comes to separating the green and yellow labels. The separation between these two is less noticeable in the other PCA cases. For k-Means, it is most likely more advantageous to use the 2D case as it seems to transform just as well as the 3D.

HTRU2

For both cases PCA seems to be a better separator than k-Means. For both cases, k-Means does not separate the data in any meaningful way, and it does not appear that one is better than the other, so if used, 2D would be the best option. Overall, PCA 3D clearly shows the best separation and is the optimal case to use for this dataset.

Digits

Both PCA and k-Means do not classify this data well, but PCA at least shows some separation. For k-Means, there is some essential data that has been taken out so it cannot be classified unless more dimensions are used. Between those two cases though, the 3D is a better option. Overall, the 3D PCA is the best option for this dataset but none of them are very good.

c. Iris

0.590901060383026

This NMI is fairly high because the data is arranged in a way that clusters can be clearly grouped together. In the graphs the data is fairly clearly separated by distance, which corresponds with the high NMI value.

HTRU2

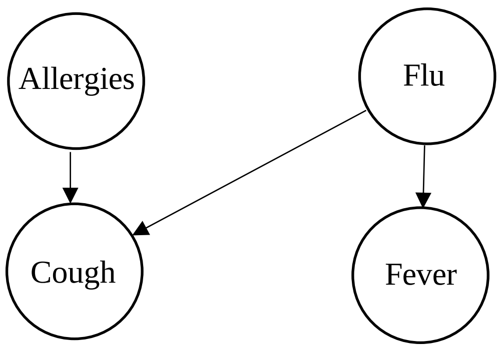
0.06791885643418222

This NMI is very low since the data is not best separated by distance. The k-Means graphs both show the data being stretched over a long distance yet the thing that distinguishes the two labels is not this distance, meaning that k-Means is not using the most important dimensions.

Digits
0.4055863492356023
This NMI is in the middle, indicating that k-Means does a decent job of clustering the data but the boundaries of each cluster are not particularly clear. This can be seen in the graphs above as certain labels seem to cover particular areas of the graph, yet the labels still overlap many times.

3. *Bayes Nets*

a.



b. $p(Cgh, Fev, Alg, Flu) = p(Cgh \mid Alg, Flu) * p(Fev \mid Flu) * p(Alg) * p(Flu)$

c. Transition Model: $p(Z_t = k \mid Z_{t-1} = j, Ut) = A_{jk}$
Observation Model: $p(X_t = l \mid Z_t = j, Ut) = B_{jl}$
Prior: $p(Z_0 = j) = \pi_j$

d. Using Bayes theorem, $p(X1, X2 \mid Y) = \frac{p(Y \mid X1, X2)p(X1, X2)}{p(Y)}$:

	Y=1	Y=0
P(X1=1, X2=1 | Y)	0	.5
P(X1=1, X2=0 | Y)	.5	0
P(X1=0, X2=1 | Y)	.5	0
P(X1=0, X2=0 | Y)	0	.5

Using Bayes theorem, $p(X1 \mid Y) = \frac{p(Y \mid X1)p(X1)}{p(Y)}$:

	Y=1	Y=0
P(X1=1 | Y)	.5	.5
P(X1=0 | Y)	.5	.5

Using Bayes theorem, $p(X2 \mid Y) = \frac{p(Y \mid X2)p(X2)}{p(Y)}$:

	Y=1	Y=0
P(X2=1 | Y)	.5	.5
P(X2=0 | Y)	.5	.5

Therefore, the following can be calculated for $p(X1 \mid Y) * p(X2 \mid Y)$:

	Y=1	Y=0
P(X1=1 | Y) * P(X2=1 | Y)	.25	.25
P(X1=1 | Y) * P(X2=0 | Y)	.25	.25
P(X1=0 | Y) * P(X2=1 | Y)	.25	.25
P(X1=0 | Y) * P(X2=0 | Y)	.25	.25

It is clear from the above tables that $p(X1, X2 \mid Y) \neq p(X1 \mid Y) * p(X2 \mid Y)$.