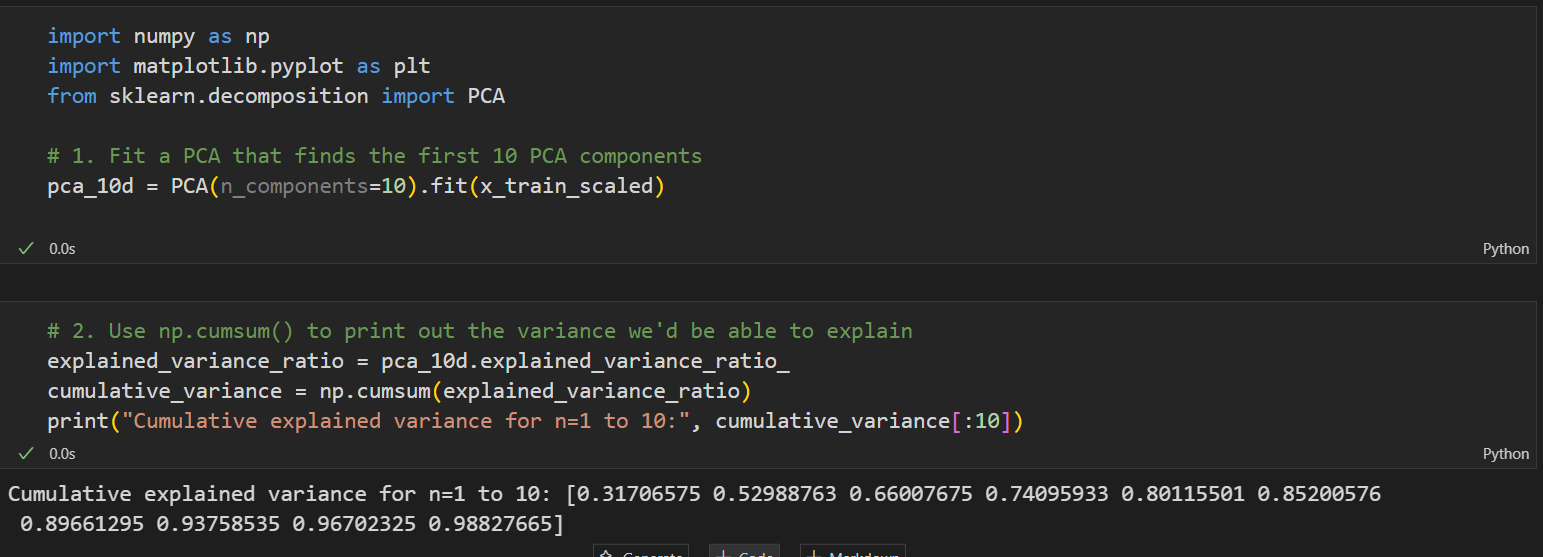
Name – Withanage W.I.N.

Index- 220731M

Assignment- PCA Machine Learning Lab

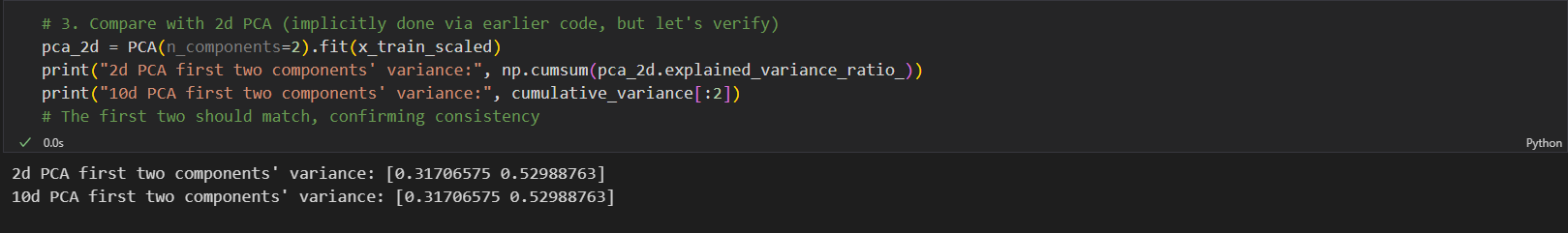
1. Fit a PCA that finds the first 10 PCA components of our training data



1. 2. Use 'np.cumsum()' to print out the variance we'd be able to explain by using n PCA dimensions for n=1 through 10

|  |  |
| --- | --- |
| Number of Component | Cumulative Variance |
| 1 | 0.31706575 |
| 2 | 0.52988763 |
| 3 | 0.66007675 |
| 4 | 0.74095933 |
| 5 | 0.80115501 |
| 6 | 0.85200576 |
| 7 | 0.89661295 |
| 8 | 0.93758535 |
| 9 | 0.96702325 |
| 10 | 0.98827665 |

1. Does the 10-dimension PCA agree with the 2d PCA on how much variance the first components explain? \*\*Do the 10d and 2d PCAs find the same first two dimensions? Why or why not?



**Agreement on variance**: Yes, the 10-dimension PCA should match the 2d PCA’s explained variance for the first two components, as both use the same data and PCA algorithm, with the 2d PCA being a subset of the 10d PCA.

**Same first two dimensions**: Yes, the first two dimensions should be identical, as PCA computes components sequentially based on maximum variance, and reducing to 2d or 10d doesn’t alter the initial components unless the data or scaling changes.

**Why**: The consistency arises because PCA’s component calculation is deterministic given fixed input (scaled data), and higher dimensions extend rather than replace earlier ones.

1. Make a plot of number of PCA dimensions against total variance explained. What PCA dimension looks good to you?

