Kernel PCA is stochastic in nature

- There is no concept of finding the best principal component.
- Resulting transformation will be different each time

→ run a classification algorithm along with kernel PCA & measure performance. Choose transformation which results in best performance

Process:

→ Mean Center data

→ kernel: $e^{-(\gamma * \text{distance}^2)}$

- Calc pairwise distances
  - results in square matrix → Symmetrix Matrix
    - think of this as a similarity matrix

→ Now find eigen decomposition of this new matrix

→ pick desired Principal components & transform data.

But where is the stochastic nature?