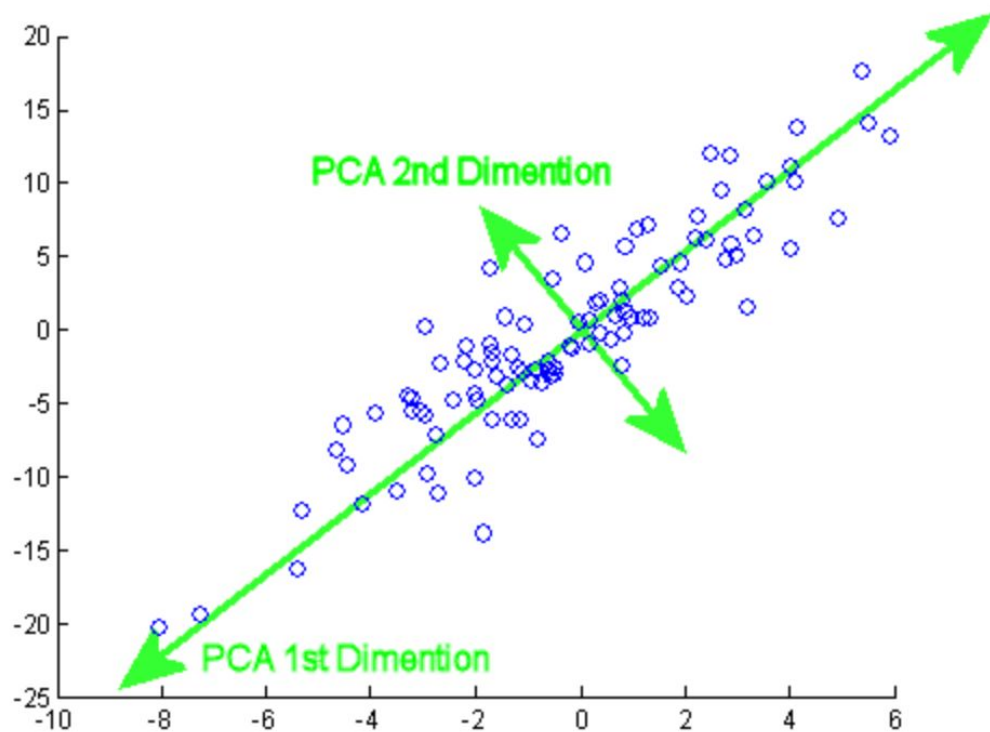


L09

PCA

Demo



HW

- Get dataset from Kaggle (any tabular dataset you want)
- Make simple classifier / regressor on the dataset
- Reasonably reduce dataset dimensionality
 - Plot explained variance
 - Explain chosen number of components
- Retrain the same classifier / regressor on the dataset with reduced dimensionality
- Compare accuracies / MSEs and speed of the two approaches (with and without dimensionality reduction)

Where get ideas for data sets:

- <https://www.meganstodel.com/posts/no-to-iris/>
- <https://machinelearningmastery.com/standard-machine-learning-datasets/>
- <https://scikit-learn.org/stable/modules/classes.html#loaders>