Unsupervised Learning Laboratories

Alex Rodriguez, Francesco Tomba

Fall Semester 2025

Intrinsic Dimension and Density Estimation

You can use external libraries for linear algebra operations but you are expected to write your own algorithms.

Exercise 1

Using the `dry_beans_dataset` as we did in previous laboratories (ie. follow the same preprocessing steps but **do not** perform a train-test split), program your own implementation of the two-NN estimate for the Intrinsic Dimension.

Is the result compatible with what you would expect from an analysis of PCA's spectrum?

Exercise 2

Using the following code, create a one-dimensional dataset of size $N = 100$.

```
X = np.concatenate(
(np.random.standard_t(1, int(0.04*N))-3.5,
np.random.normal(5, 1, int(0.48 * N)),
np.random.normal(7.5, 1, int(0.48 * N))
))[:, np.newaxis]
```

Compute the density estimation with your implementations of:

- Histogram Density Estimation (Freedman Diaconis rule)
- Kernel Density Estimation (KDE) - Gaussian kernel (Silverman's rule)