Project2_JL lemma

Johnson-Lindenstrauss Lemma

Any high dimensional dataset can be randomly projected into a lower dimensional Euclidean space while controlling the distortion in the pairwise distances.

∃ linear transformation $\Phi : R^d \rightarrow R^k$ $(k \ll d)$ s.t

- $\forall\, i, j \quad (1-\epsilon)\|\mathbf{v_i} - \mathbf{v_j}\|_2 \leq \|\Phi(\mathbf{v_i}) - \Phi(\mathbf{v_j})\|_2$
- $\leq (1+\epsilon)\|\mathbf{v_i} - \mathbf{v_j}\|_2$
- where $k \leq \frac{c \ln n}{\epsilon^2}$ and $\epsilon \in \left(0, \frac{1}{2}\right)$

Reference : FODS23S lec12 Dimension Reduction v1

Program question

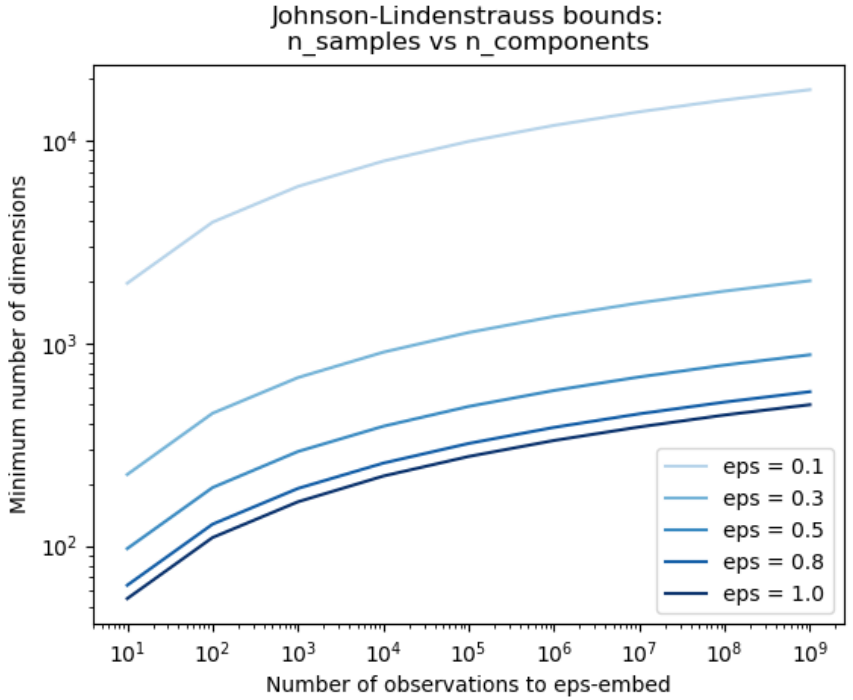
- Where the code (***JL_Lemma.ipynb***) is **None** (as shown below),

  please fill in the correct code.

  (The code ***JL_Lemma.ipynb*** will be uploaded to ecourse2)

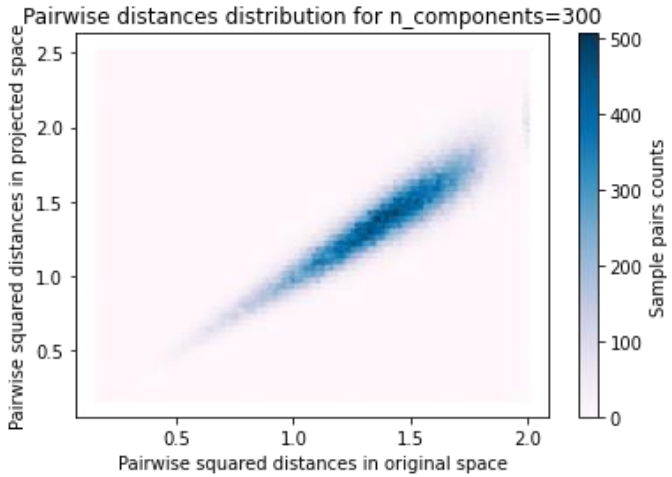
Experiment 1

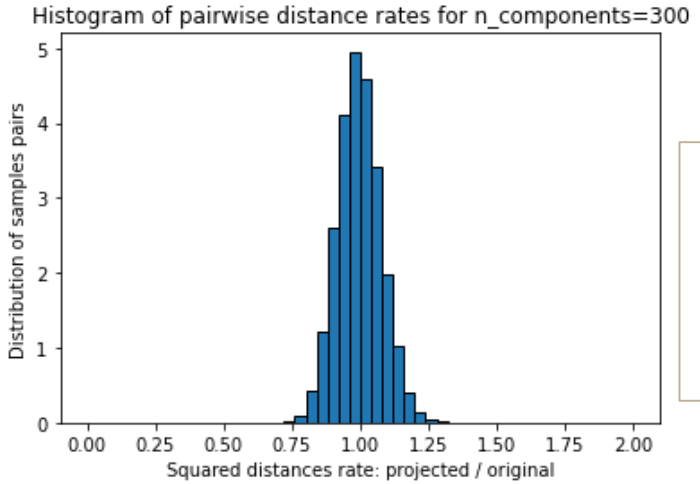
- Find the relationship between **the number of samples (x_axis)** and the **minimum number of dimensions (y_axis)** to show the impact of different **distortion bounds**.



Experiment 2

- Use **Gaussian (N(0,1)) Random projection** to construct matrix Φ.
- Take the first 500 samples from the 20 newsgroups dataset, and observe the degree of distortion of **reducing the dimension from 130107 to [300, 1000, 10000]**

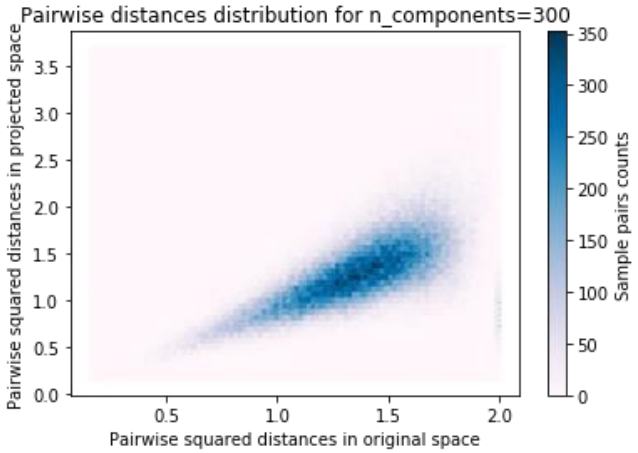


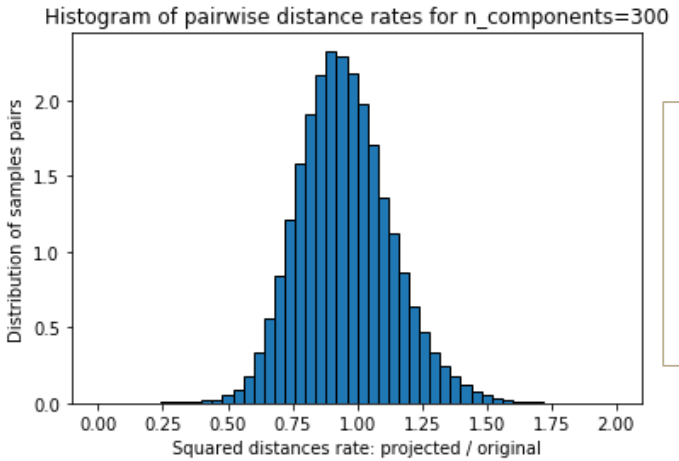



```
Embedding 500 samples with dim 130107 using various random projections
Projected 500 samples from 130107 to 300 in 1.296s
Mean distances rate: 0.99 (0.08)
Projected 500 samples from 130107 to 1000 in 4.041s
Mean distances rate: 1.00 (0.04)
Projected 500 samples from 130107 to 10000 in 960.713s
Mean distances rate: 1.00 (0.01)
```

Experiment 3

- Use **Sparse ([+1, -1]) Random projection (lec12_v1 pp.31)** to construct matrix Φ.
- Take the first 500 samples from the 20 newsgroups dataset, and observe the degree of distortion of **reducing the dimension from 130107 to [300, 1000, 10000]**






```
Embedding 500 samples with dim 130107 using various random projections
Projected 500 samples from 130107 to 300 in 0.305s
Mean distances rate: 0.88 (0.17)
Projected 500 samples from 130107 to 1000 in 0.922s
Mean distances rate: 1.01 (0.10)
Projected 500 samples from 130107 to 10000 in 8.995s
Mean distances rate: 0.97 (0.04)
```

Report

1. Explain what you observed from the figure generated from Experiment 1 with **distortion bound ϵ = 0.1, 0.3, 0.5, 0.8, 1.0** , respectively. (eps =ϵ) (10%)
2. Explain what you observed from Experiment 2. (10%)
3. Compare the results of Experiment 2 and Experiment 3 with two properties given in lec12_v1 pp.17, and explain the difference. (10%)

Submit file

- **Code** : StudentNumber_Name.**ipynb**
- **Report** : StudentNumber_Name**.pdf**

Deadline : 5/6 (23:59)