Comparison of Intrinsic Dimension Estimators

Aim: Compare the performance of different ID estimators. We will evaluate these estimators based on their mean and standard deviation.

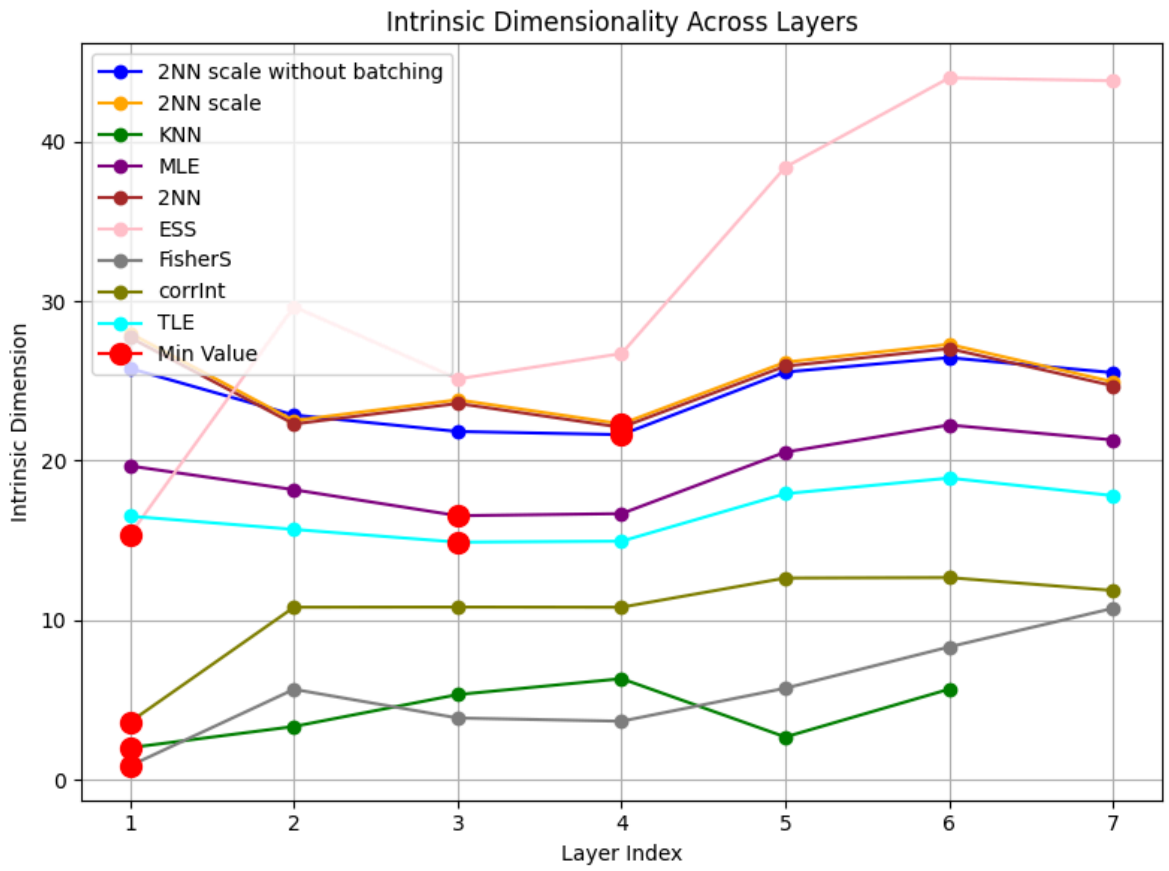
Experiment 1:

Experiments Setup:
Model: DistilBert
Dataset: Multiclass Sentiment Analysis Dataset
Number of samples: 1000
Batches: 3
Evaluation Method: Mean Intrinsic Dimension and Standard Deviation (Calculated from the intrinsic dimensions obtained in each batch).

Results:

Intrinsic Dimension Means across 3 batches

Layer ID	1	2	3	4	5	6	7
2NN no batch	25.78	22.86	21.83	21.63	25.56	26.47	25.52
2NN scale	27.99	22.52	23.82	22.31	26.18	27.3	24.95
KNN	2.	3.33	5.33	6.33	2.67	5.67	258.67
MLE	19.67	18.18	16.55	16.68	20.54	22.24	21.3
2NN	27.72	22.3	23.59	22.09	25.93	27.03	24.7
ESS	15.38	29.66	25.13	26.72	38.41	44.01	43.83
FisherS	0.87	5.66	3.86	3.66	5.73	8.32	10.74
CorrInt	3.62	10.81	10.82	10.81	12.64	12.67	11.86
TLE	16.52	15.69	14.89	14.95	17.93	18.91	17.81



Approaches:

1. **Intrinsic Dimension estimation using 2NN Algorithm with Scaling on 1000 samples without batching**
    - Method: ‘return_id_scaling_2NN’ [source]

Layer ID	Intrinsic Dimension
1	25.783684711530256
2	22.862326971615232
3	21.83180717509225
4	21.63306150660067
5	25.56115256440055
6	26.465031360283753
7	25.521369064933594

2. **Intrinsic Dimension estimation using 2NN Algorithm with Scaling**
    - Method: ‘return_id_scaling_2NN’ [source]

Approach:
Intrinsic dimensionality is estimated using the 2NN algorithm across various scales, from the entire dataset down to smaller subsets. For each layer in the transformer, we select the ID associated with the scale with the minimum error, ensuring the most accurate representation of the data's dimensionality.

In this experiment, the intrinsic dimension at the first scale, which is the full subset, is considered. At this scale, the error is always zero.

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	27.99	0.83
2	22.52	0.66
3	23.82	2.13
4	22.31	0.16
5	26.18	2.33
6	27.3	1.71
7	24.95	2.71

3. **Intrinsic Dimension estimation using kNN algorithm**
    - Method: ‘skdim.id.KNN’ [source]

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	2.	0.82
2	3.33	1.89
3	5.33	4.71
4	6.33	4.78
5	2.67	0.94
6	5.67	2.87
7	258.67	360.15

4. **Intrinsic dimension estimation using the Maximum Likelihood algorithm**
    - Method: ‘skdim.id.MLE’ [source]

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	19.67	0.47
2	18.18	0.32
3	16.55	0.39
4	16.68	0.09
5	20.54	0.48
6	22.24	0.72
7	21.3	0.7

5. **Intrinsic dimension estimation using the TwoNN algorithm.**
    - Method: ‘skdim.id.TwoNN’ [source]

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	27.72	0.83
2	22.3	0.66
3	23.59	2.11
4	22.09	0.16
5	25.93	2.31
6	27.03	1.7
7	24.7	2.69

6. **Intrinsic dimension estimation using the Expected Simplex Skewness algorithm.**
    - Method: ‘skdim.id.ESS’ [source]

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	15.38	0.95
2	29.66	0.65
3	25.13	0.56
4	26.72	0.5
5	38.41	1.02
6	44.01	1.44
7	43.83	0.46

7. **Intrinsic dimension estimation using the Fisher Separability algorithm.**
    - Method: ‘skdim.id.FisherS’ [source]

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	0.87	0.01
2	5.66	0.71
3	3.86	0.36
4	3.66	0.03
5	5.73	0.32
6	8.32	0.83
7	10.74	0.57

8. **Intrinsic dimension estimation using the Correlation Dimension.**
    - Method: ‘skdim.id.CorrInt’ [source]

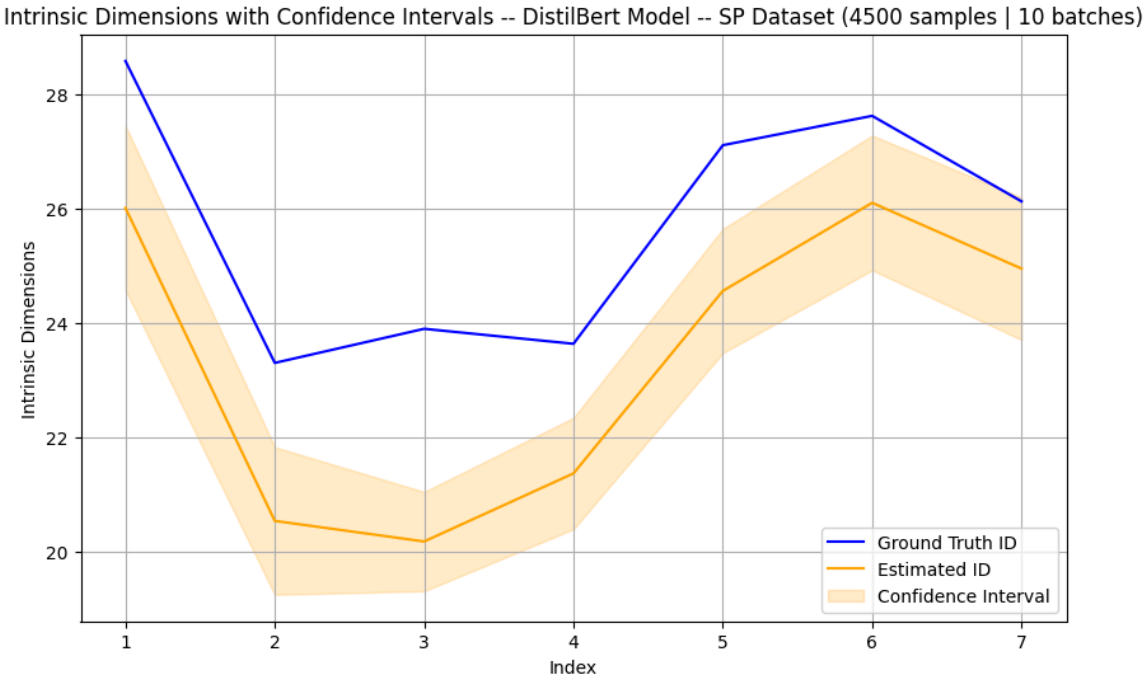
Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	3.62	0.79
2	10.81	0.34
3	10.82	0.2
4	10.81	0.27
5	12.64	0.62
6	12.67	0.83
7	11.86	0.16

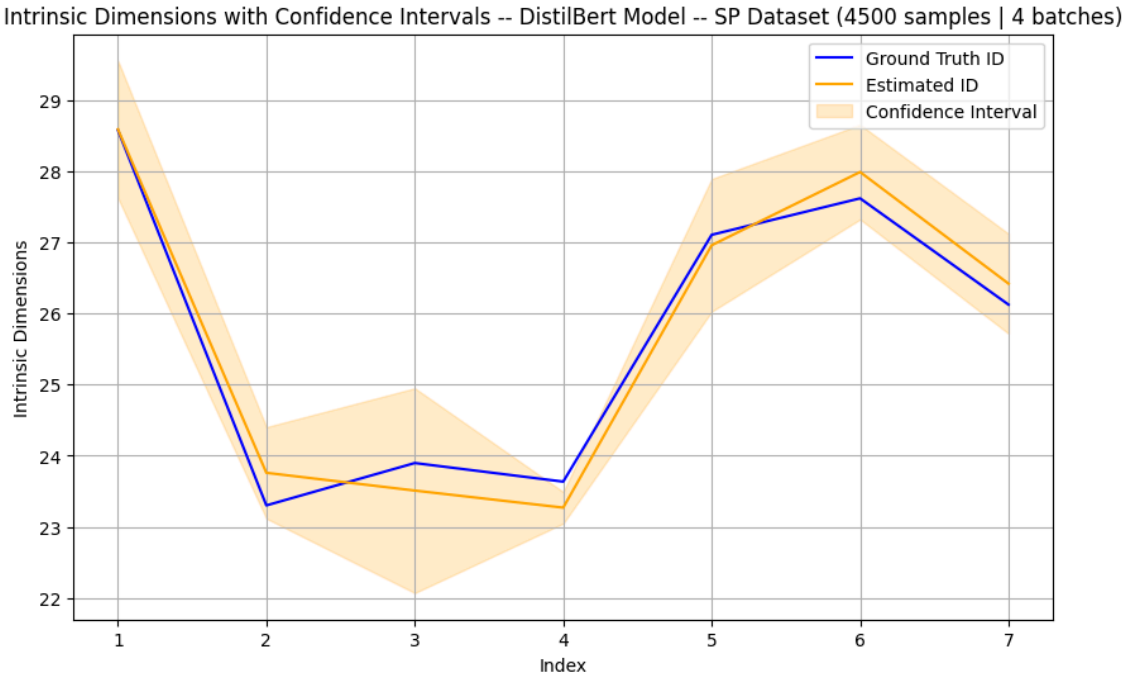
9. **Intrinsic dimension estimation using the Tight Local intrinsic dimensionality Estimator algorithm.**
    - Method: ‘skdim.id.TLE’ [source]

Layer ID	Mean Intrinsic Dimension	Standard Deviation
1	16.52	0.32
2	15.69	0.26
3	14.89	0.32
4	14.95	0.11
5	17.93	0.42
6	18.91	0.63
7	17.81	0.48

Results: Intrinsic Dimensions with Confidence Intervals

Ground Truth ID - Intrinsic Dimension estimation using 2NN Algorithm with scaling
Estimated ID - Intrinsic Dimension estimation using 2NN Algorithm with scaling averaged (weighted) over number of samples



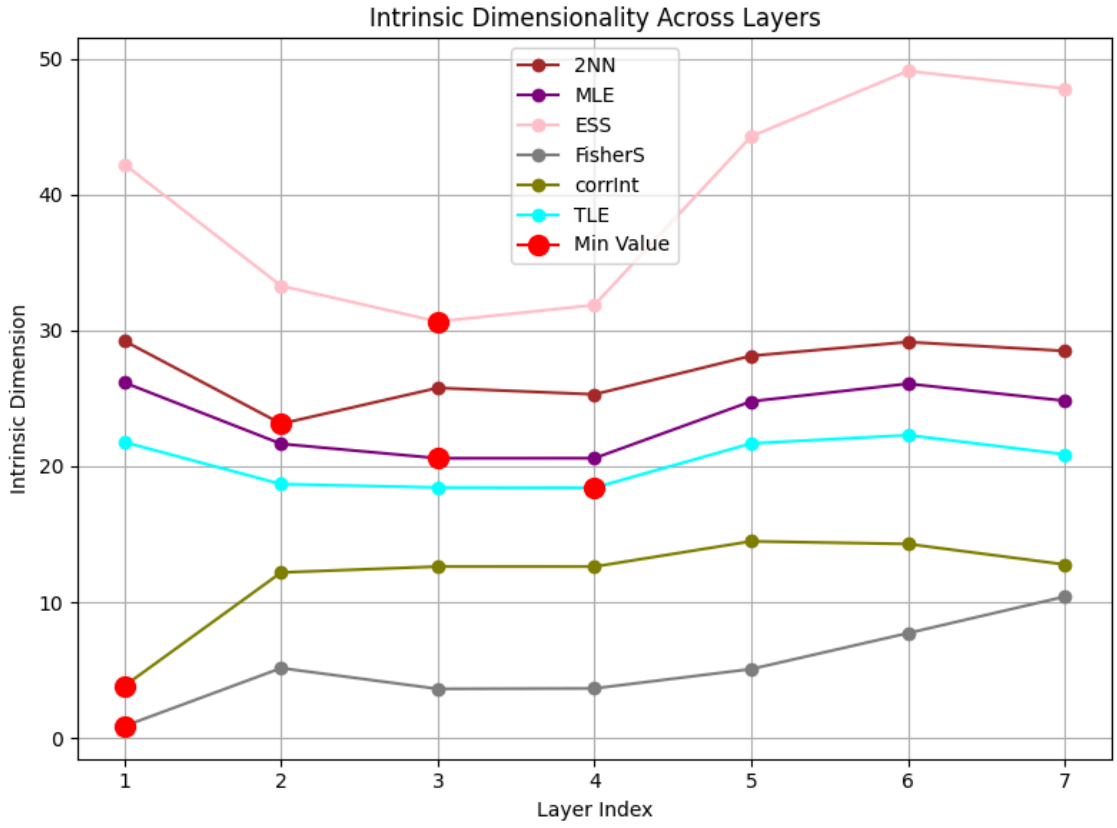


Analysis of 10 Experimental Trials

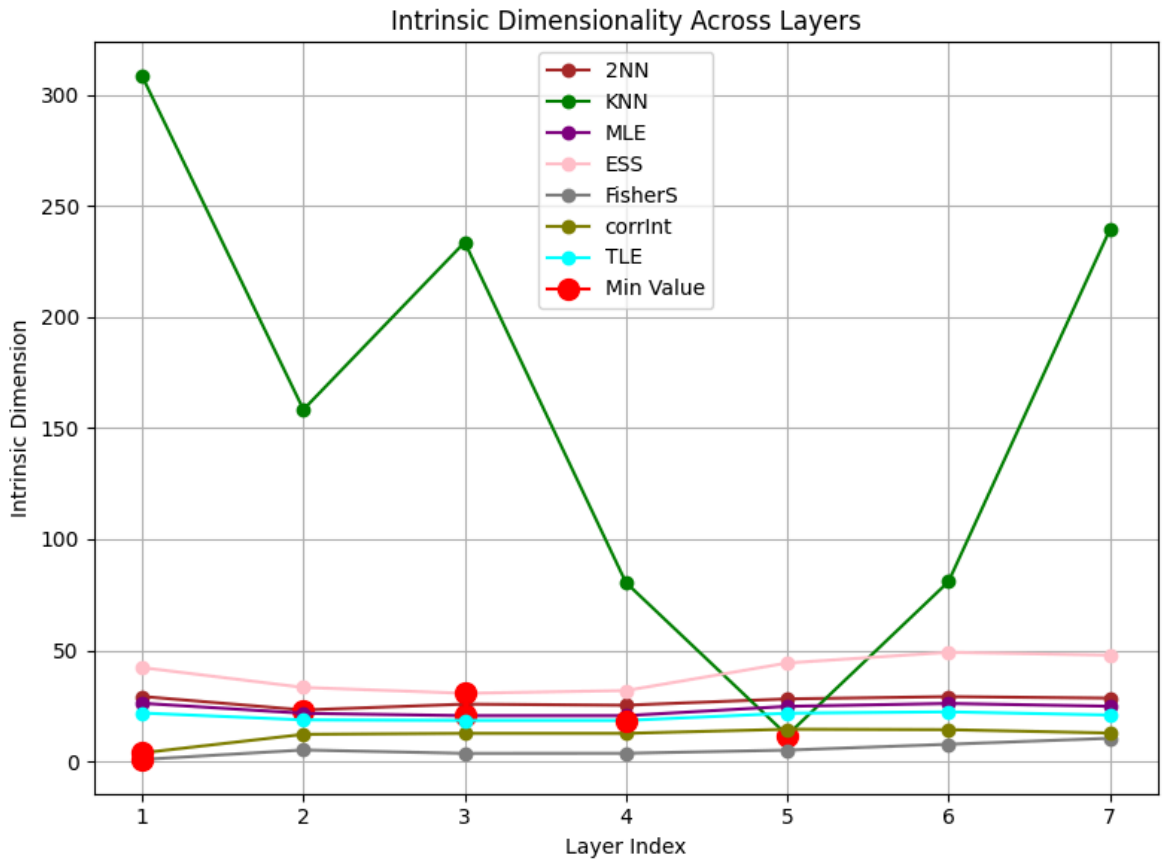
The table below presents the mean intrinsic dimensions and standard deviations across 10 trial runs, using various ID estimators.

Layer ID	1	2	3	4	5	6	7
2NN	29.23 ± 0	23.14 ± 0	25.78 ± 0	25.30 ± 0	28.14 ± 0	29.16 ± 0	28.50 ± 0
KNN	308.4 ± 375.26	158.4 ± 304.82	233.7 ± 349.78	80.5 ± 229.18	11.8 ± 15.75	80.9 ± 229.04	239.2 ± 346.24
MLE	26.17 ±0	21.66 ±0	20.60 ±0	20.60 ±0	24.79 ±0	26.08 ±0	24.83 ±0
FisherS	0.87 ±0	5.14 ±0	3.61 ±0	3.65 ±0	5.07 ±0	7.72 ±0	10.40 ±0
CorrInt	3.83 ±0	12.19 ±0	12.62 ±0	12.62 ±0	14.48 ±0	14.28 ±0	12.78 ±0
TLE	21.78 ±0	18.70 ±0	18.44 ±0	18.43 ±0	21.68 ±0	22.30 ±0	20.88 ±0
ESS	42.23 ±0	33.29 ±0	30.66 ±0	31.88 ±0	44.30 ±0	49.11 ±0	47.82 ± 0

Mean ID vs Layers for different ID Estimators (without KNN estimator results)

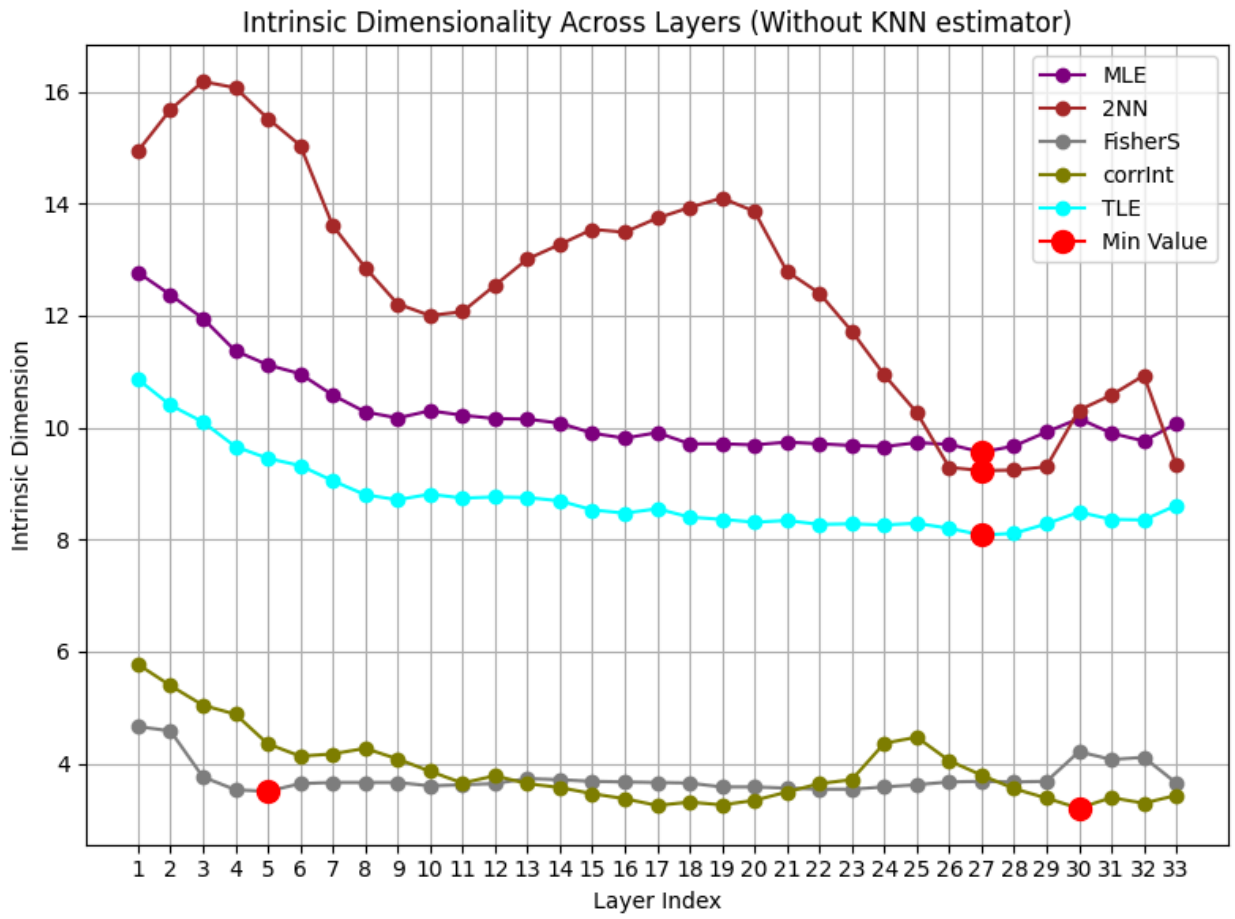


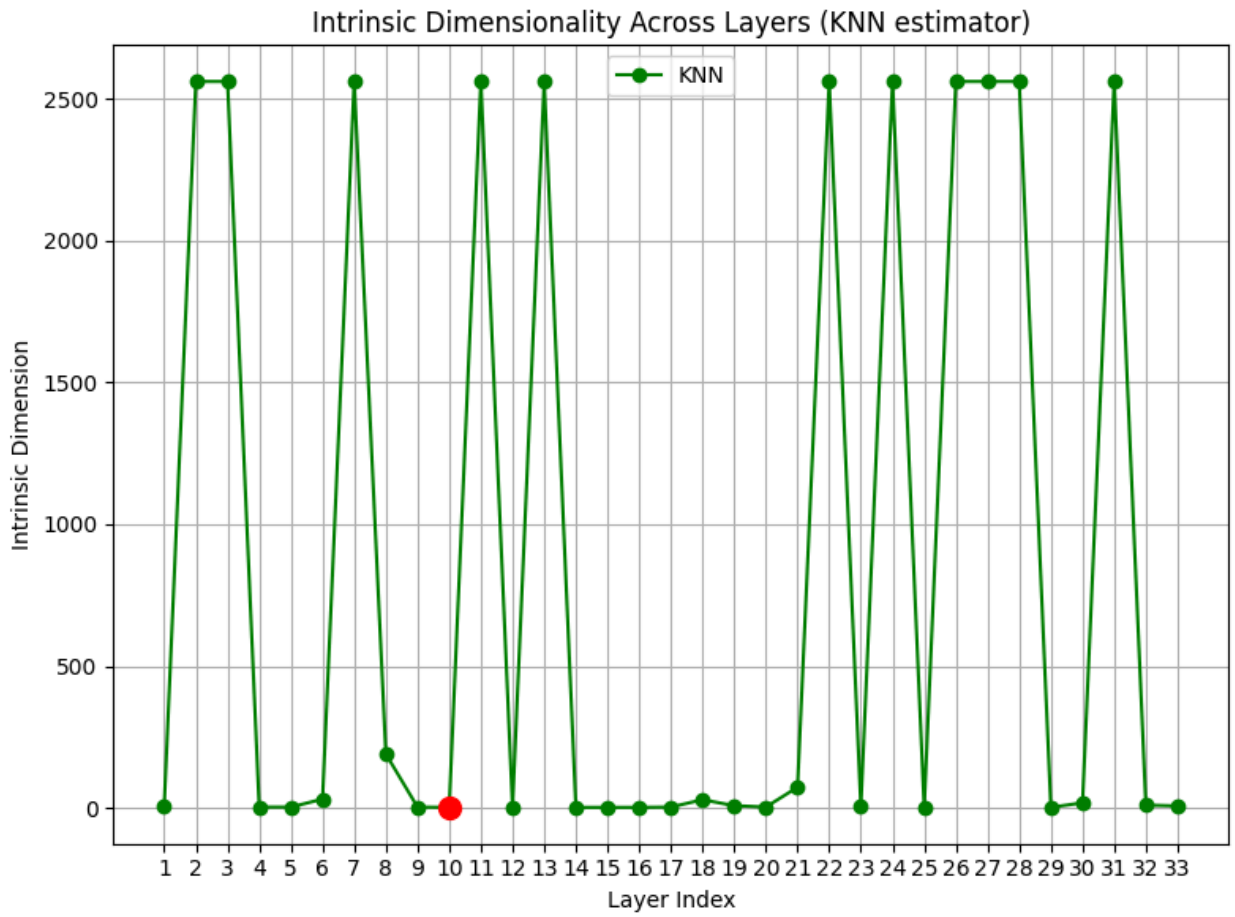
Mean ID vs Layers for different ID Estimators (with KNN estimator results)



Experiment 2:

Experiments Setup:
Model: Phi2
Dataset: MedQuad
Number of samples: 200
Batches: 25





1. **Intrinsic dimension estimation using the TwoNN algorithm.**
    - Method: 'skdim.id.TwoNN' [source]

Intrinsic dimensions for each of the 33 layers of the Phi2 model:

14.94, 15.68, 16.18, 16.07, 15.52, 15.04, 13.61, 12.86, 12.21, 12.0, 12.07, 12.54, 13.01, 13.27, 13.54, 13.49, 13.74, 13.93, 14.1, 13.86, 12.79, 12.4, 11.72, 10.94, 10.27, 9.29, 9.23, 9.24, 9.3, 10.31, 10.58, 10.93, 9.33

2. **Intrinsic Dimension estimation using kNN algorithm**
    - Method: 'skdim.id.KNN' [source]

Intrinsic dimensions for each of the 33 layers of the Phi2 model:

6, 2560, 2560, 3, 3, 31, 2560, 190, 3, 2, 2560, 3, 2560, 2, 2, 2, 3, 29, 8, 3, 73, 2560, 4, 2560, 2, 2560, 2560, 2560, 2, 18, 2560, 10, 7

3. **Intrinsic dimension estimation using the Maximum Likelihood algorithm**
    - Method: ‘skdim.id.MLE’ [source]

Intrinsic dimensions for each of the 33 layers of the Phi2 model:

12.77, 12.37, 11.95, 11.37, 11.12, 10.96, 10.58, 10.28, 10.17, 10.3, 10.22, 10.16, 10.15, 10.08, 9.9, 9.81, 9.91, 9.71, 9.71, 9.69, 9.74, 9.71, 9.68, 9.66, 9.73, 9.7, 9.56, 9.67, 9.92, 10.16, 9.9, 9.76, 10.07

4. **Intrinsic dimension estimation using the Fisher Separability algorithm.**
    - Method: ‘skdim.id.FisherS’ [source]

Intrinsic dimensions for each of the 33 layers of the Phi2 model:

4.66, 4.58, 3.76, 3.53, 3.5, 3.64, 3.66, 3.66, 3.66, 3.6, 3.62, 3.64, 3.73, 3.71, 3.68, 3.67, 3.66, 3.65, 3.58, 3.58, 3.56, 3.54, 3.54, 3.58, 3.62, 3.67, 3.68, 3.67, 3.68, 4.21, 4.07, 4.11, 3.65

5. **Intrinsic dimension estimation using the Correlation Dimension.**
    - Method: ‘skdim.id.CorrInt’ [source]

Intrinsic dimensions for each of the 33 layers of the Phi2 model:

5.77, 5.39, 5.04, 4.88, 4.35, 4.13, 4.17, 4.27, 4.08, 3.86, 3.65, 3.78, 3.64, 3.57, 3.46, 3.37, 3.25, 3.31, 3.26, 3.34, 3.49, 3.64, 3.71, 4.36, 4.47, 4.05, 3.79, 3.55, 3.38, 3.2, 3.39, 3.29, 3.43

6. **Intrinsic dimension estimation using the Tight Local intrinsic dimensionality Estimator algorithm.**
    - Method: ‘skdim.id.TLE’ [source]

Intrinsic dimensions for each of the 33 layers of the Phi2 model:

10.86, 10.4, 10.1, 9.66, 9.45, 9.32, 9.05, 8.8, 8.71, 8.81, 8.74, 8.76, 8.75, 8.69, 8.53, 8.47, 8.55, 8.4, 8.36, 8.31, 8.34, 8.27, 8.28, 8.26, 8.29, 8.2, 8.08, 8.11, 8.28, 8.49, 8.36, 8.35, 8.61