



AIDOS LAB
AI FOR DATA-ORIENTED SCIENCE



Intrinsic dimension

Lecture 5 — Geometry and Topology in Machine Learning

Inés García Redondo — October 13, 2026

Motivation

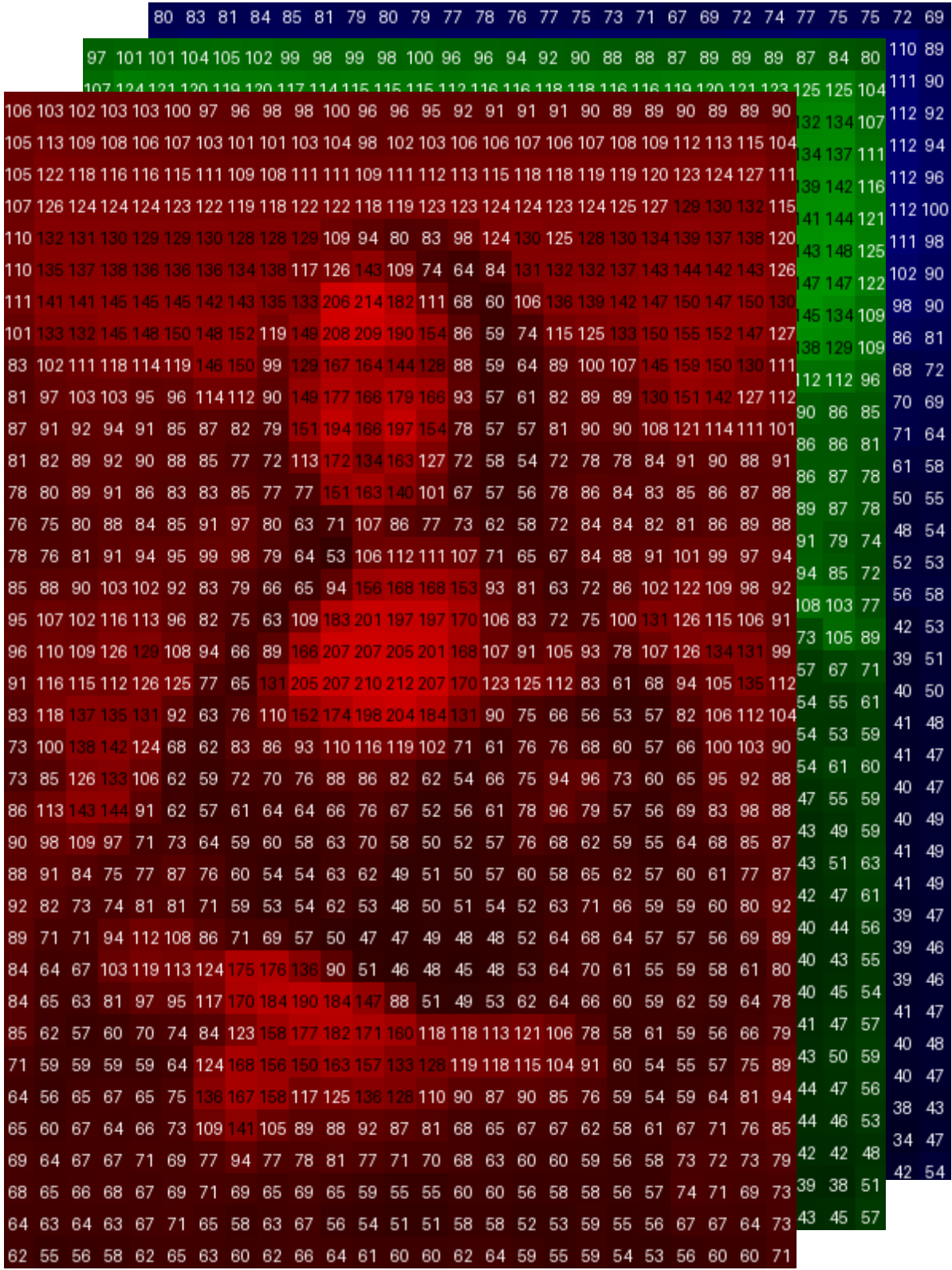
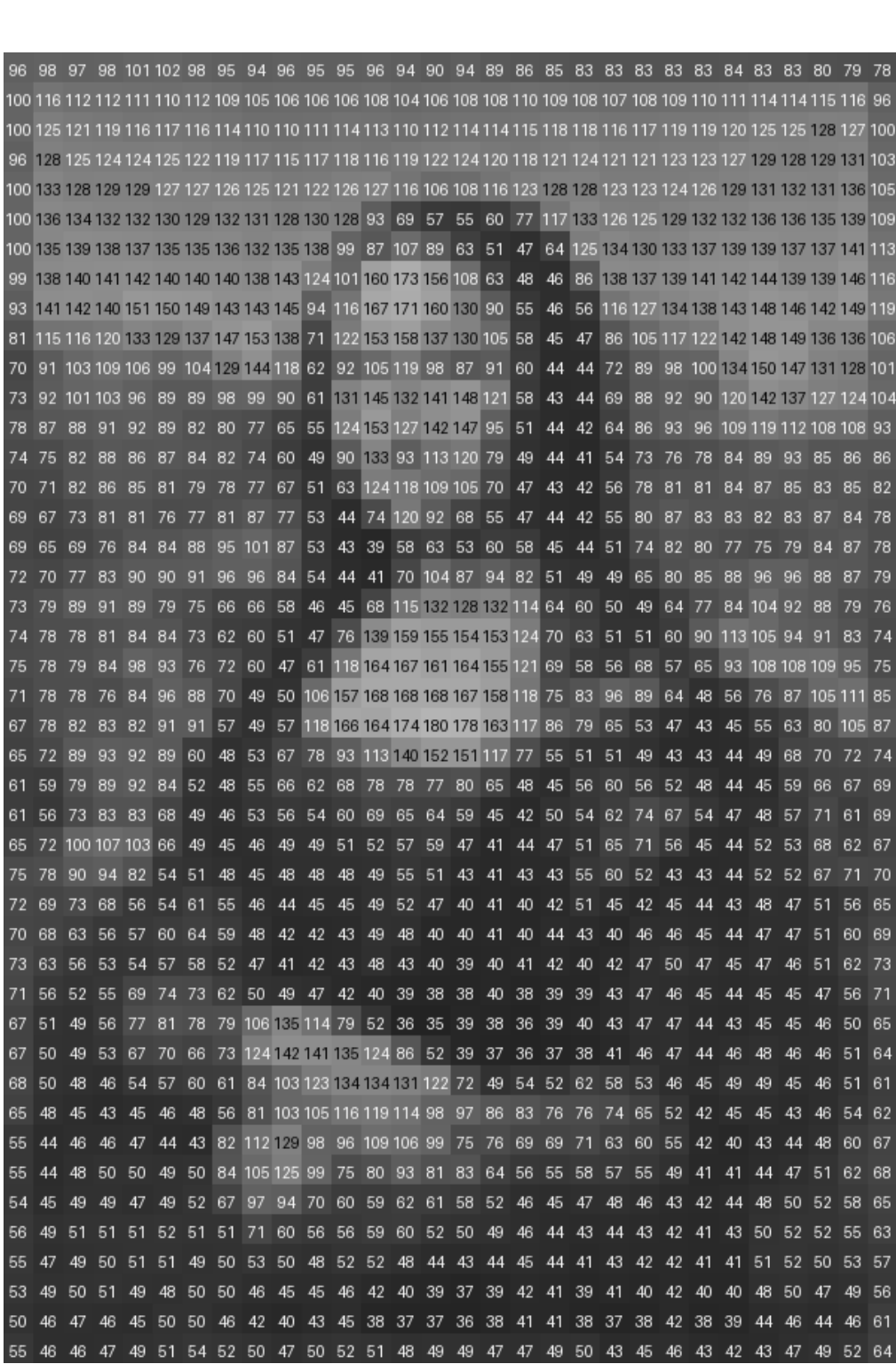
The manifold hypothesis

Typically, the data we work with is very high-dimensional

For instance, an image with resolution 512×512 lives in a 262,144 dimensional space (they are represented by matrices with that number of parameters, and if we consider them to be color images, we need to multiply the dimension by 3!)

However, meaningful images do not live at any point of that *ambient* space

Manifold hypothesis: data lives in a lower dimensional manifold embedded in the high-dimensional ambient space



$$\mathcal{M}^d \subseteq \mathbb{R}^D, \quad \dim \mathcal{M} = d \ll D$$

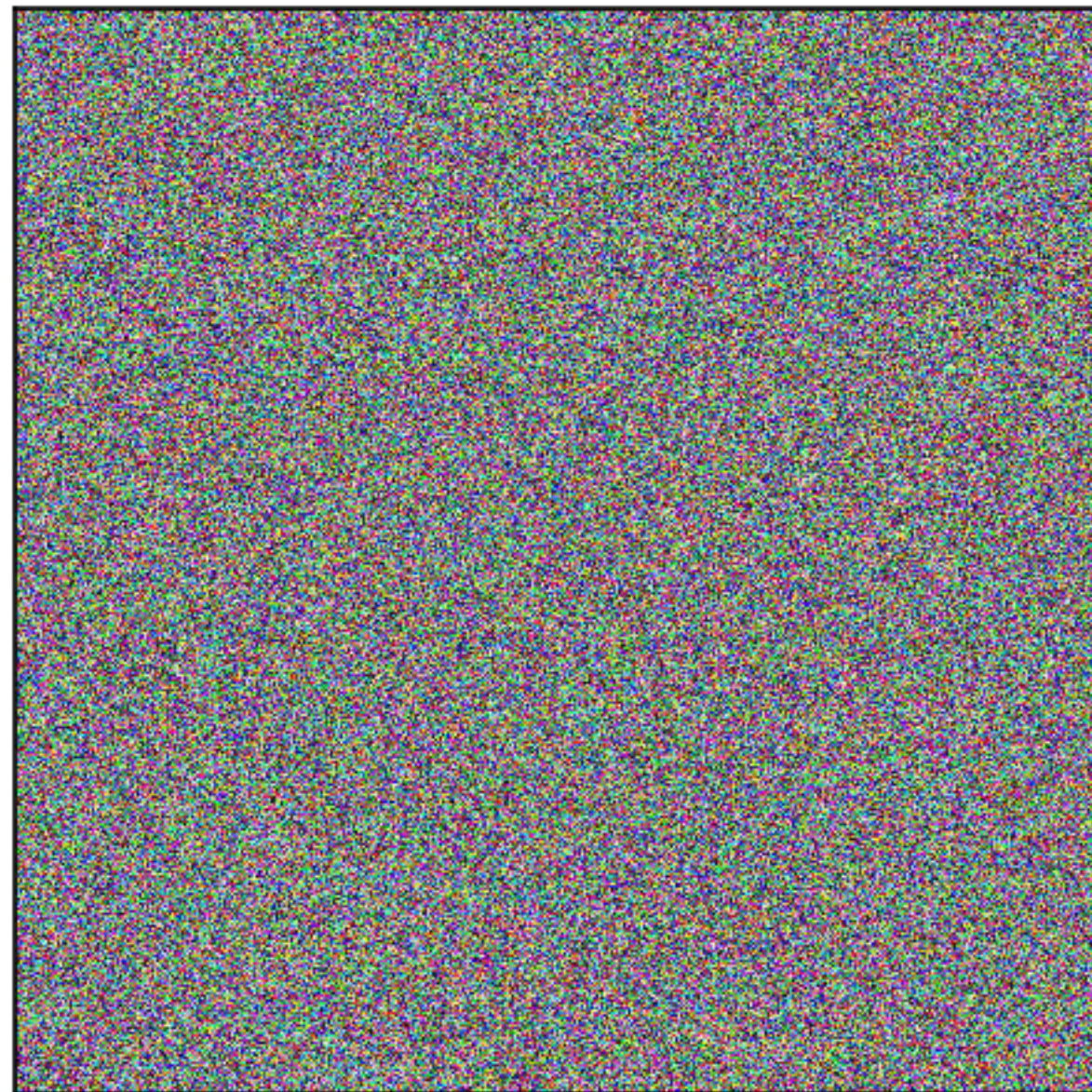
Here *manifold* will refer to a subspace that can be described with a smaller set of degrees of freedom in the ambient space

Motivation

The manifold hypothesis

- **The distribution of images is highly concentrated in this ambient space:** sampling random values at each pixel never yields a nice-looking image. Real images have very particular pixel values, where the order also matters.

Independent random RGB values



Photograph pixels shuffled



Original photograph

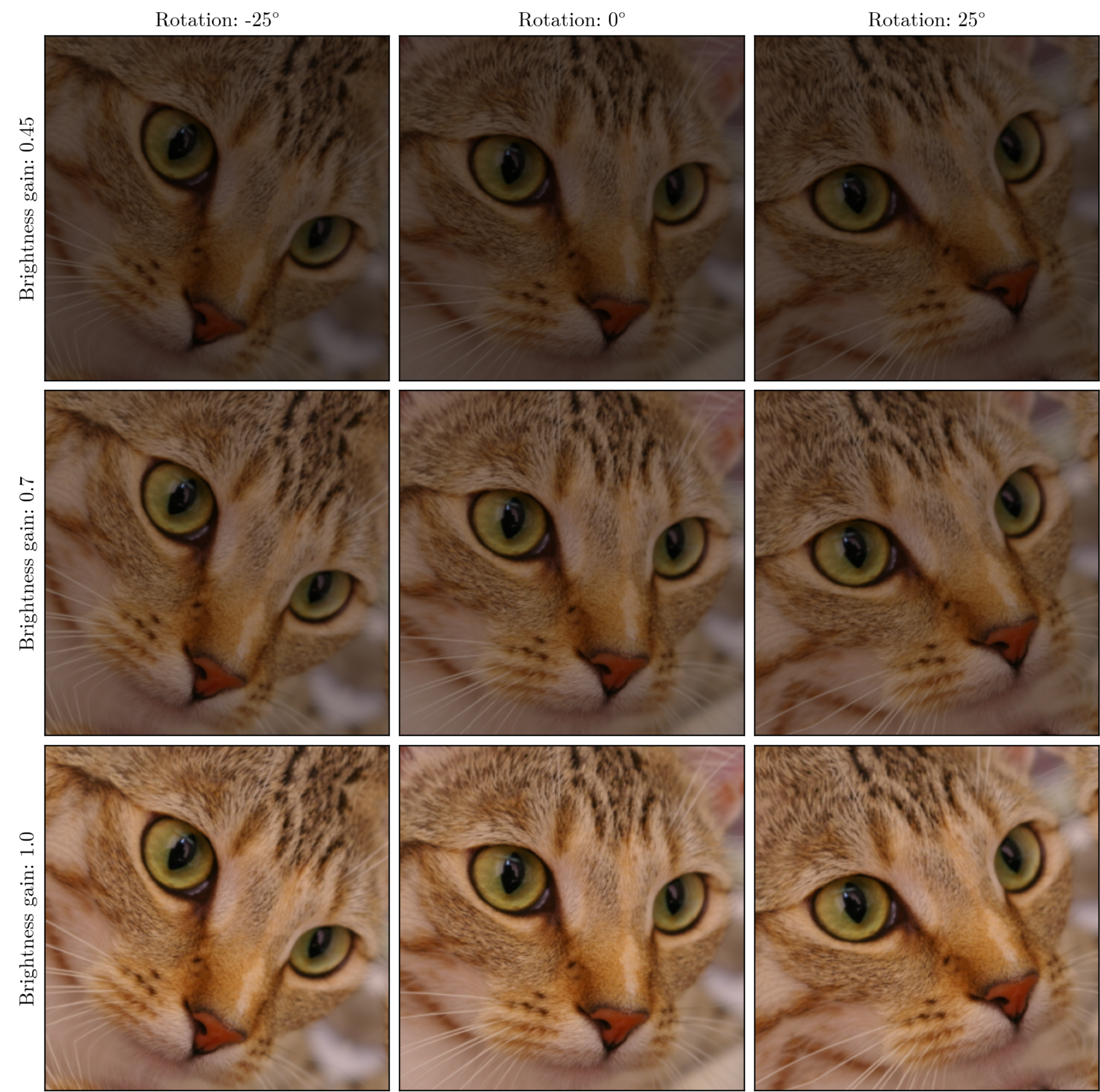


Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2, pp. 1-800). Cambridge: MIT press.

Motivation

The manifold hypothesis

- **Meaningful examples are connected to each other through continuous transformations:** we can traverse the manifold to move from one example to another

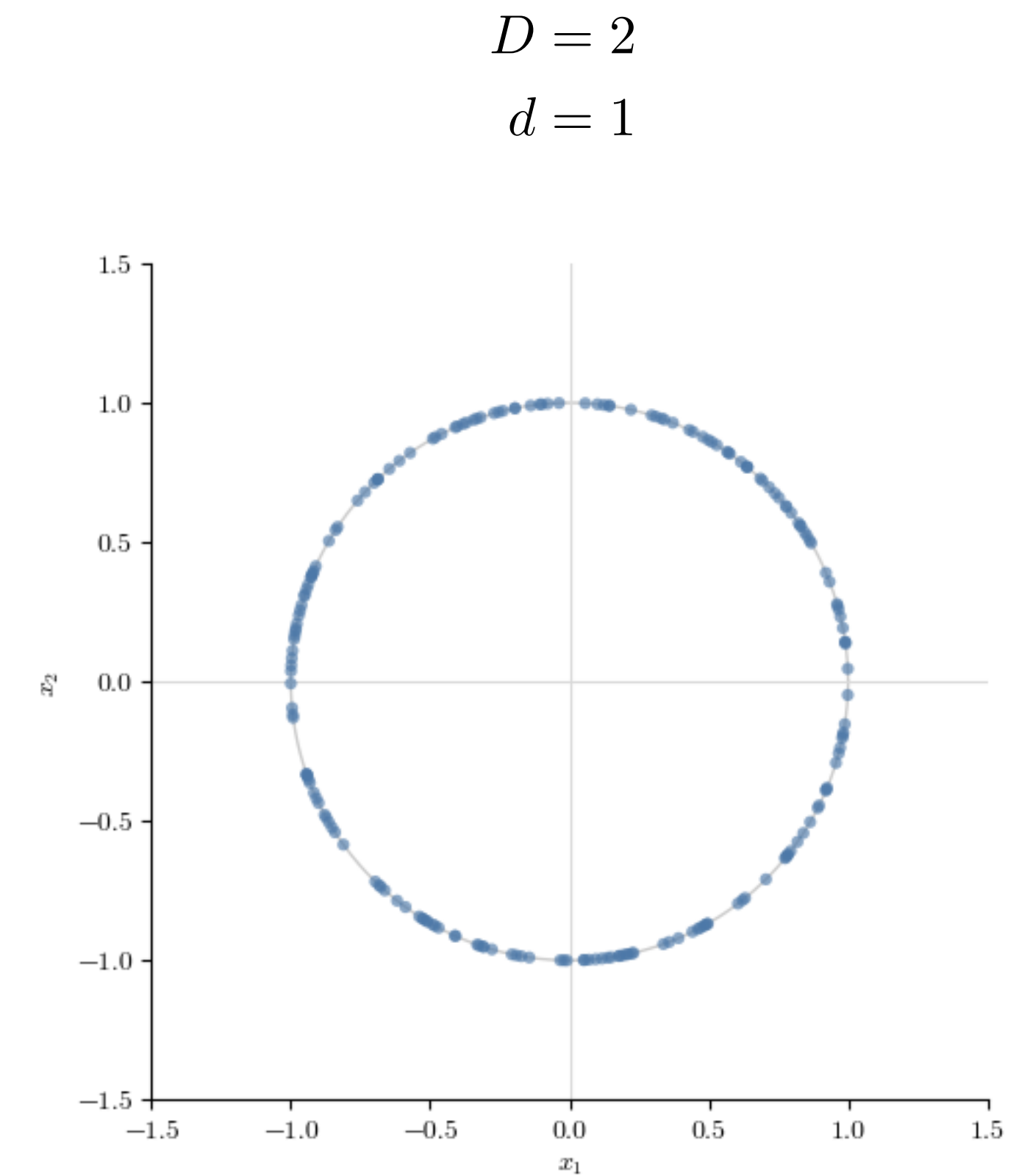


Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2, pp. 1-800). Cambridge: MIT press.

The manifold hypothesis

Relevant notations

- **Ambient dimension:** dimension D of the ambient space, the high-dimensional space where data live
- **Intrinsic dimension:** dimension d of the data manifold, or equivalently, number of local coordinates needed to describe the data
- **Estimated dimension:** a statistic $\hat{d}$ depending on samples, noise and scale



The manifold hypothesis

Relevant notations

What do we mean by manifold?

- **Local flatness:** locally, a manifold is approximately a *tangent plane*

Example: $x(\theta) = (\cos \theta, \sin \theta)$

Near a particular angle, Taylor expansion gives:

$$x(\theta_0 + h) = x(\theta_0) + h(-\sin \theta_0, \cos \theta_0) + O(h^2)$$

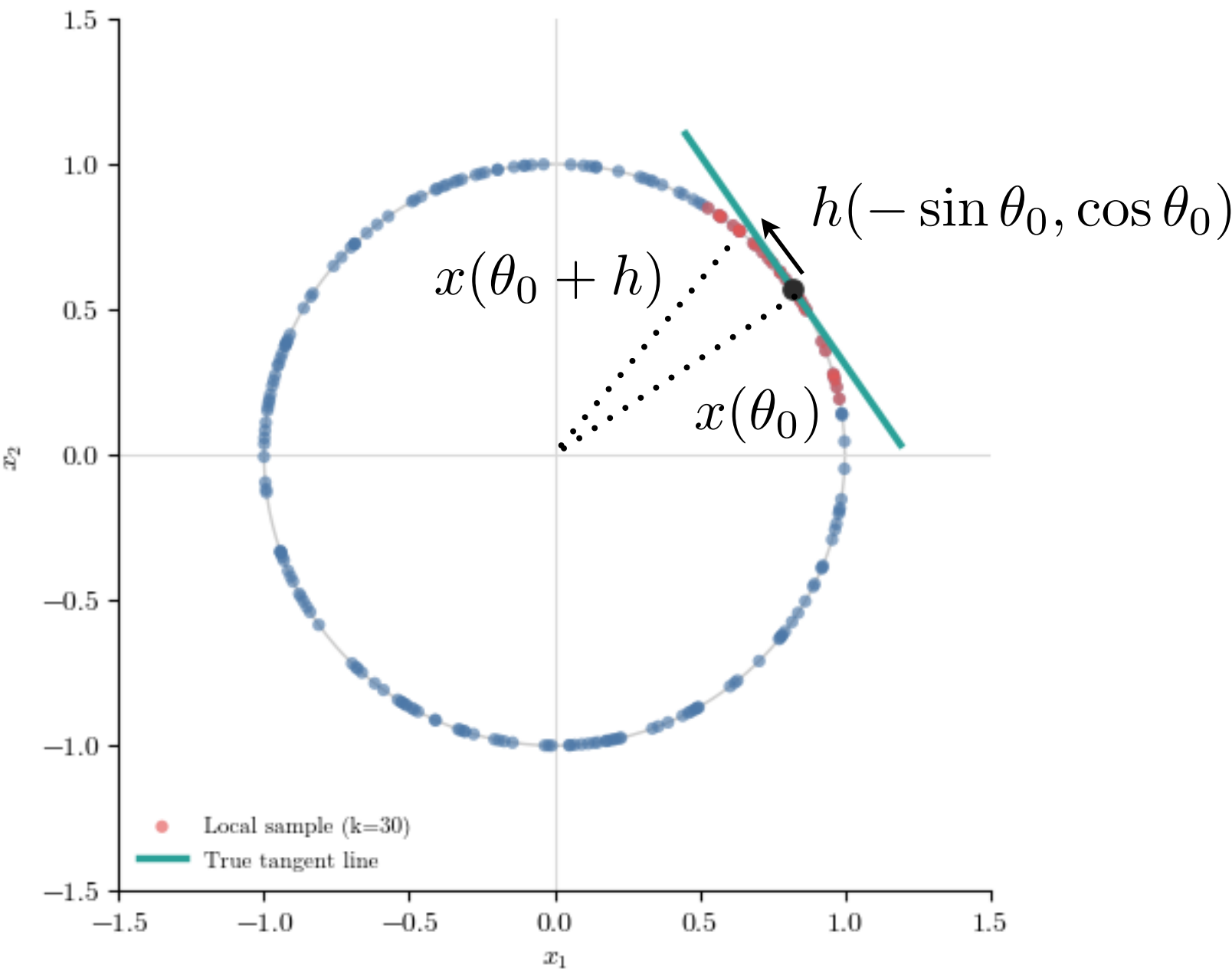
Tangent direction

Near a point $x \in \mathcal{M}$

$$\varphi(u) = x + Au + O(\|u\|^2), \quad u \in \mathbb{R}^d$$

$$A \in \mathbb{R}^{D \times d}, \quad \text{rank } A = d$$

$$D = 2$$
$$d = 1$$



The manifold hypothesis

Relevant notations

What do we mean by manifold?

■ **Local growth of volume balls:** in $\mathbb{R}^d$ the volume of a ball grows like

$$\text{vol}(B^d(x, r)) = v_d r^d, \quad v_d = \text{vol}(B(\mathbf{0}, 1))$$

Because a manifold is locally like $\mathbb{R}^d$, we also have that for small radius:

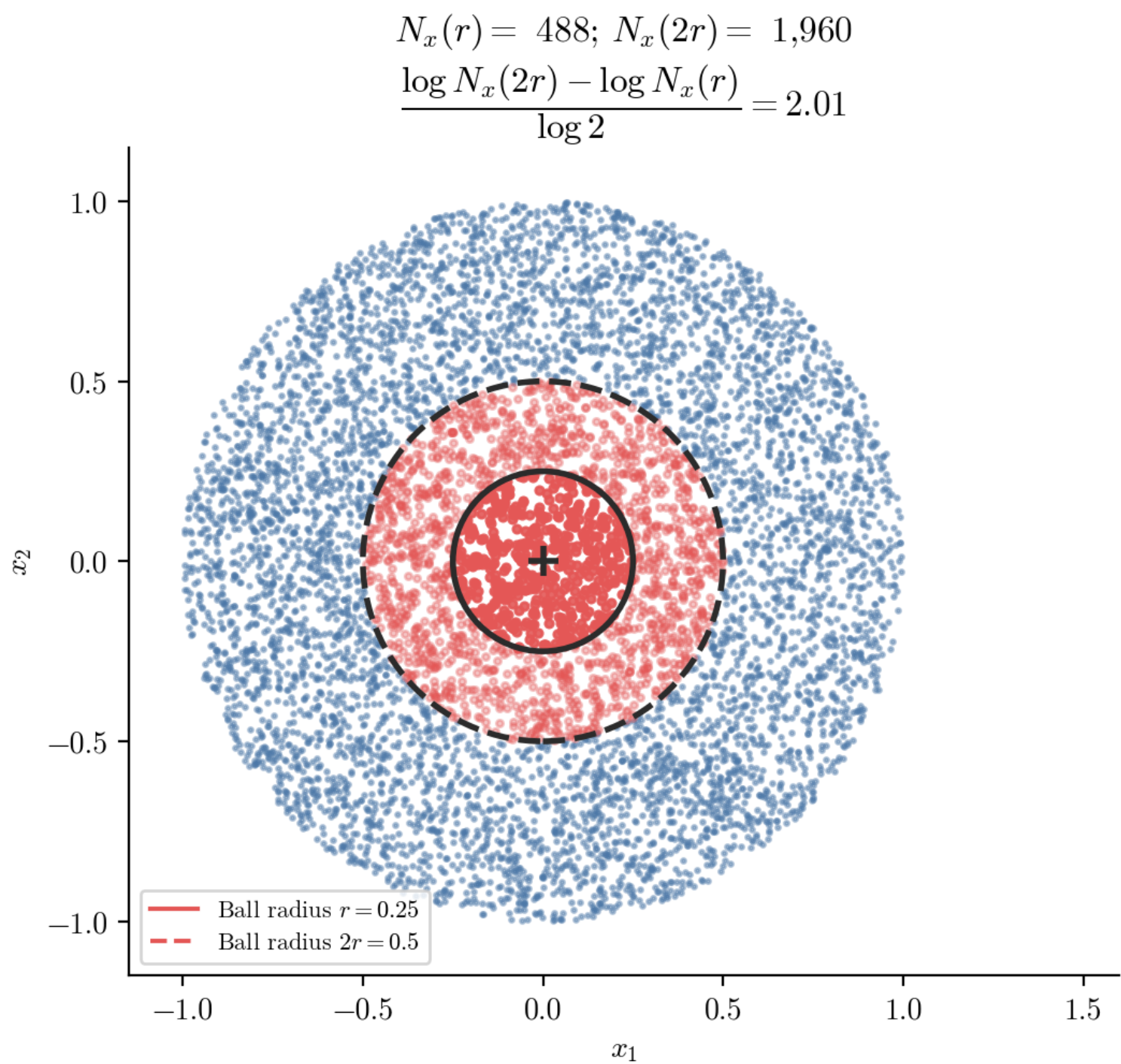
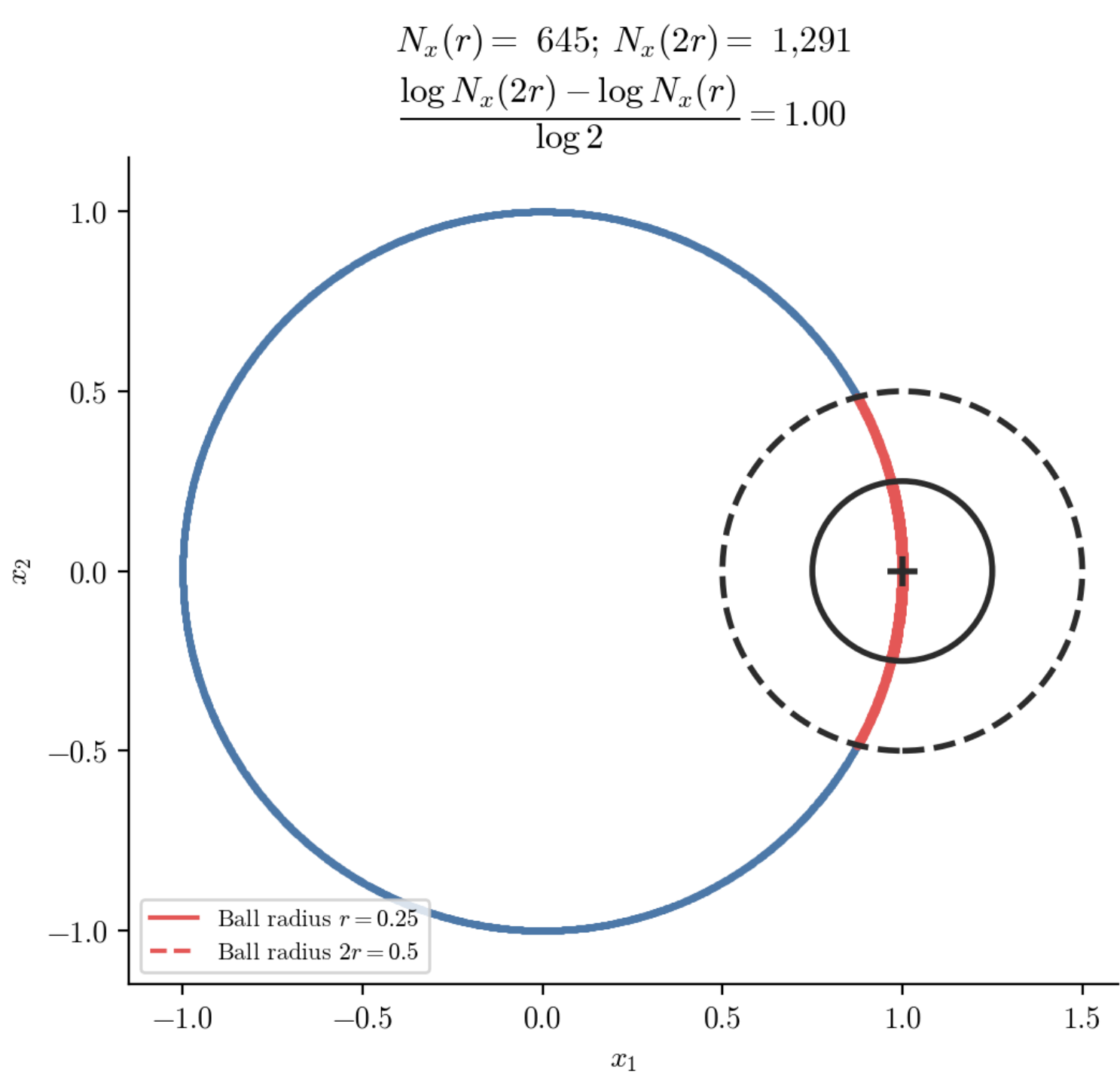
$$\text{vol}(\mathcal{M} \cap B^D(x, r)) \approx \text{vol}(B^d(x, r)) = v_d r^d$$

If we assume that we sample points from the uniform distribution over the manifold

$$\Pr(X \in \mathcal{M} \cap B^D(x, r)) \approx p(x) \text{vol}(\mathcal{M} \cap B^D(x, r)) \approx p(x) v_d r^d$$

*Proportion of sampled
points in the ball*

$$\frac{N_x(2r)}{N_x(r)} \approx \frac{p(x) v_d (2r)^d}{p(x) v_d r^d} = 2^d \implies d \approx \frac{\log N_x(2r) - \log N_x(r)}{\log 2}$$



Lecture 5 — Intrinsic dimension

Outline

- How to measure intrinsic dimension
 - Local PCA
 - TwoNN
 - MLE
- Applications of intrinsic dimension measures in deep learning

Local PCA

Lecture 5 — Intrinsic dimension

Principal component analysis

Recap

- Given m points $\mathbf{x} = \{x_1, \dots, x_m\} \subset \mathbb{R}^D$

- Take the mean: $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$

- Center all points and compute the covariance matrix: $\Sigma = \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^\top \in \mathbb{R}^{D \times D}$

- This matrix is symmetric and positive semidefinite:

$$\Sigma^\top = \left[\sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^\top \right]^\top = \sum_{i=1}^m \left[(x_i - \bar{x})(x_i - \bar{x})^\top \right]^\top = \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^\top = \Sigma$$

$$v^\top \Sigma v = \sum_{i=1}^m v^\top (x_i - \bar{x})(x_i - \bar{x})^\top v = \sum_{i=1}^m \left[(x_i - \bar{x})^\top v \right]^\top (x_i - \bar{x})^\top v = \sum_{i=1}^m \left\| (x_i - \bar{x})^\top v \right\|^2 \geq 0$$

- So we can diagonalize it (i.e., find D eigenvalues and eigenvectors): $\Sigma v_i = \lambda_i v_i, \quad 0 \leq \lambda_D \leq \lambda_{D-1} \leq \dots \leq \lambda_1$

The eigenvectors (also called **principal components**) capture the directions in which the data changes the most

The first k principal components capture the following fraction of the total variation:

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\sum_{j=1}^D \lambda_j}$$

Principal component analysis

A linear subspace

If our point cloud lies exactly on a d -dimensional affine plane,
we can express our sample as

$$x_i = a + Uz_i, \quad U \in \mathbb{R}^{D \times d}, \text{rank}(U) = d \text{ and } z_i \in \mathbb{R}^d$$

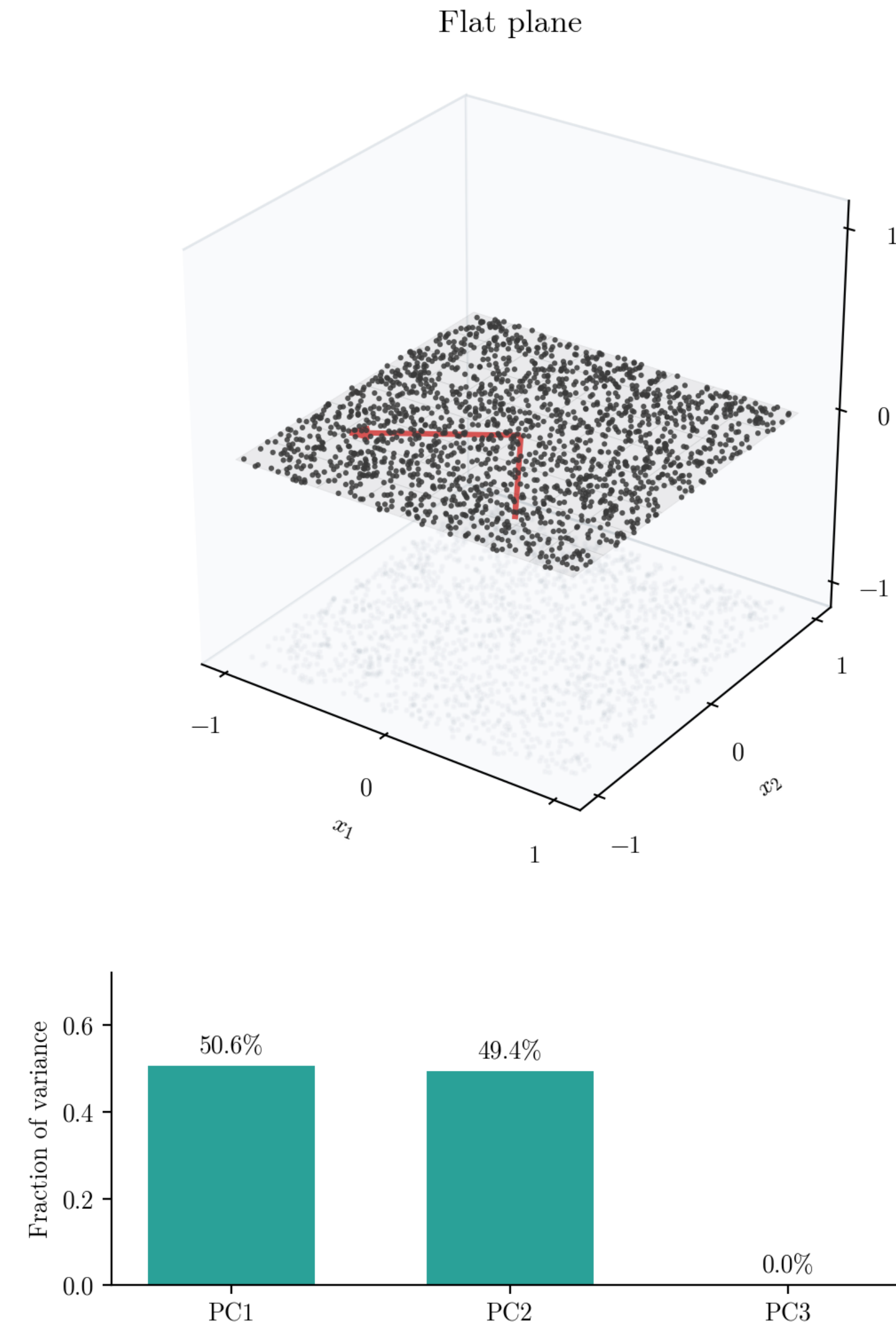
Centering the points, we get:

$$x_i - \bar{x} = Uz_i - U \left[\frac{1}{m} \sum_{i=1}^m z_i \right] = U(z_i - \bar{z})$$

If the z_i range $\mathbb{R}^d$, we conclude:

$$\Sigma_x = U \Sigma_z U^\top \implies \text{rank}(\Sigma_x) = d$$

Or in other words, the **covariance matrix of the PCA will have exactly d non-zero eigenvalues**, expressing these independent directions of variation



Local PCA

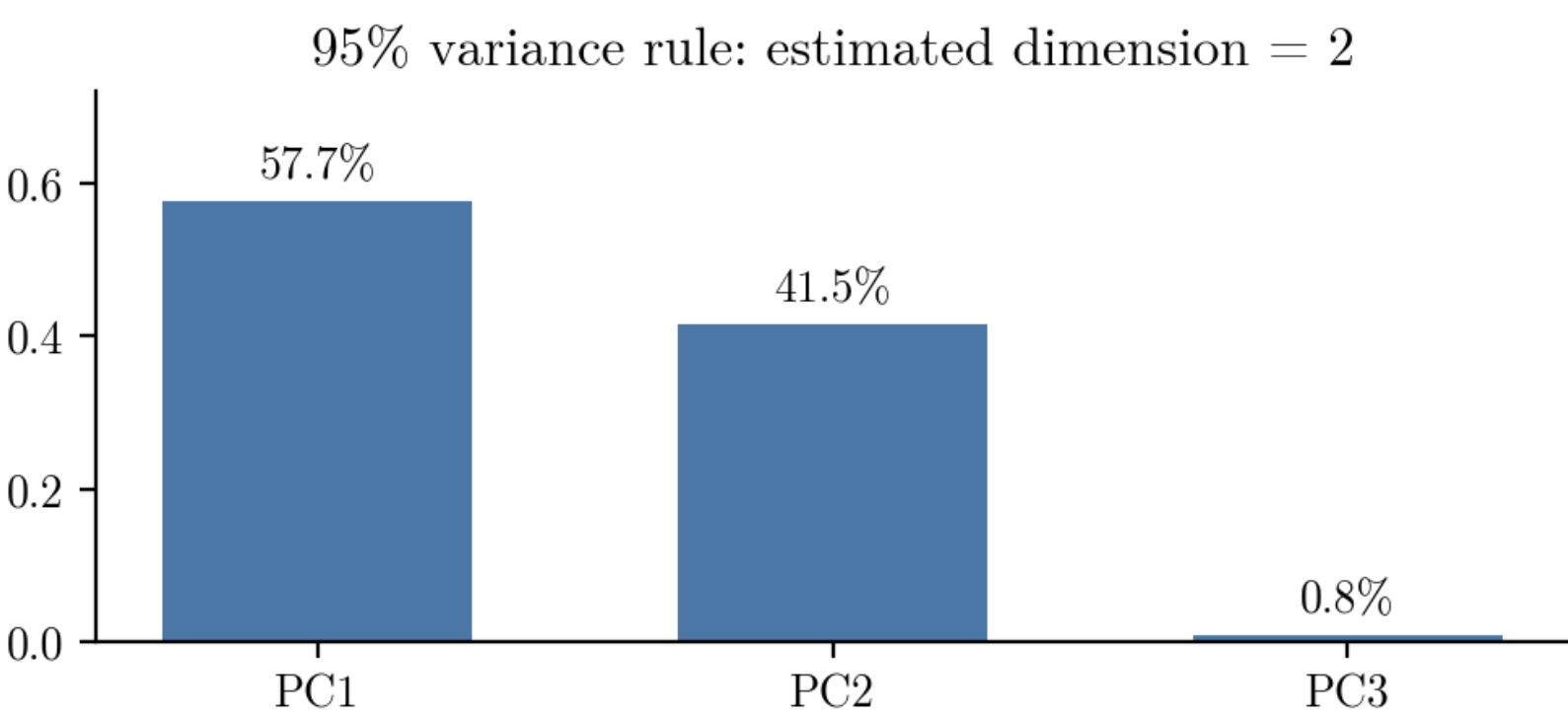
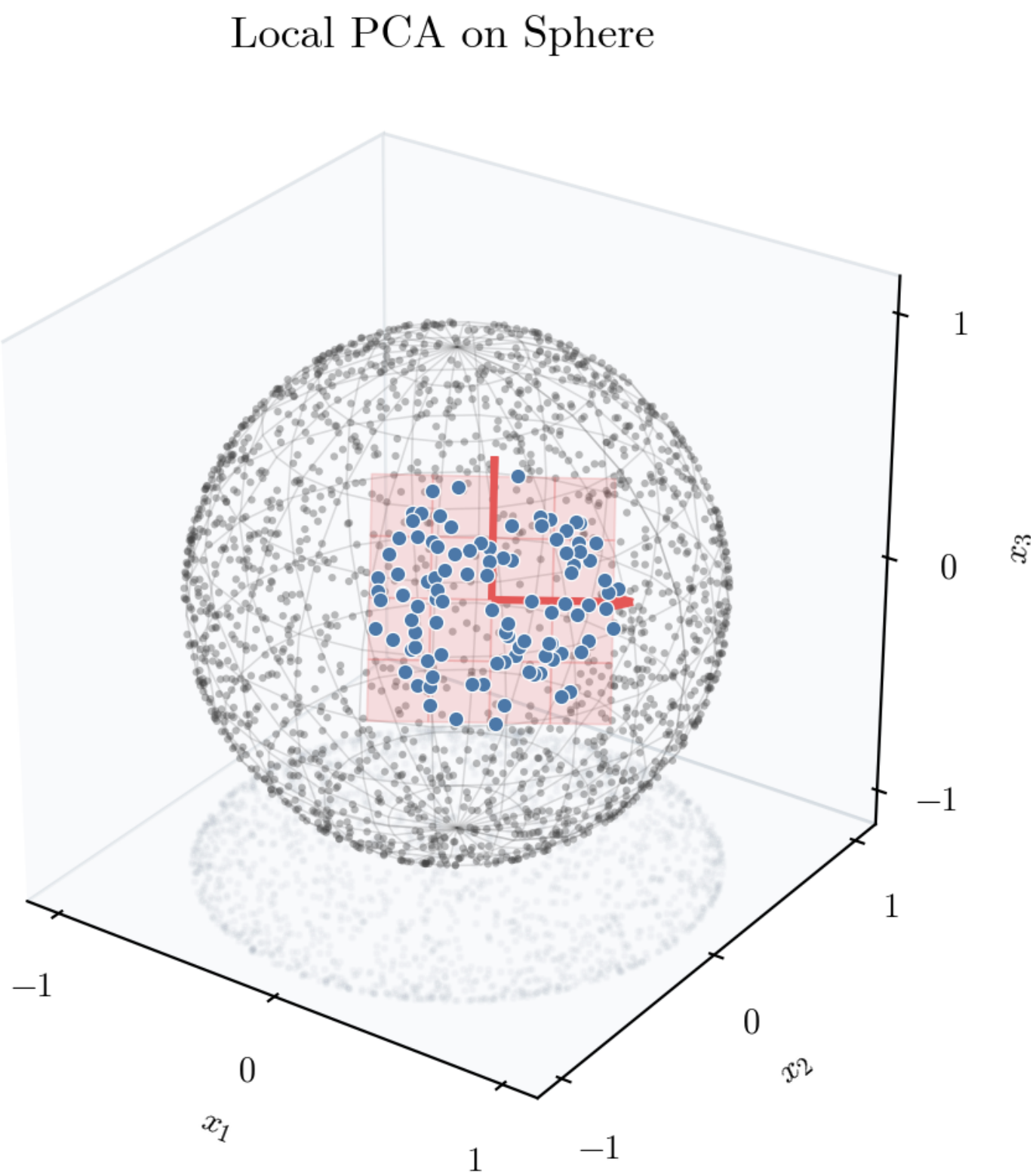
Approximating the dimension of the manifold

So if our points lie on a manifold, which locally is like its tangent plane, then we can estimate the dimension around a point in our sample by **applying PCA to a neighborhood around the point**

- Pick a point in your sample x
- Select the k nearest neighbors (k is a parameter of the method)
- Apply PCA to the point cloud formed by x and the neighbors

If we stay in a small neighborhood, we will find that a few eigenvalues explain most of the variation, corresponding to the linear directions of the tangent space around the points

Setting a threshold (e.g. 95% of the variation), we can estimate the dimension around x as the number of eigenvalues needed to explain at least that amount of variation



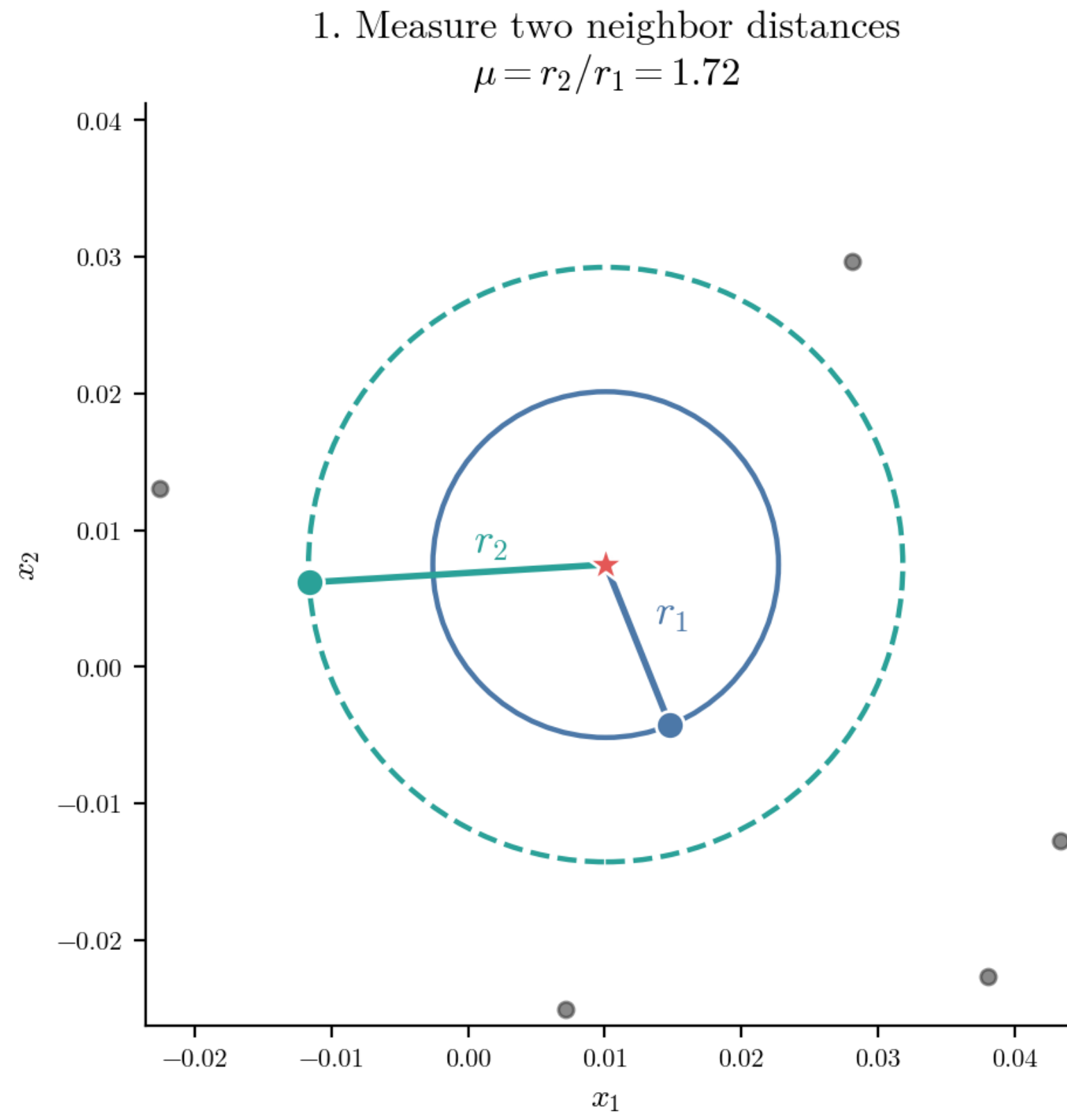
Lim, U., Oberhauser, H., and Nanda, V. (2024). Tangent space and dimension estimation with the Wasserstein distance. *SIAM Journal on Applied Algebra and Geometry*, 8(3), 650-685.

TwoNN

Lecture 5 — Intrinsic dimension

TwoNN

Approximating the dimension of the manifold



Let r_1 and r_2 be the distances to the two nearest neighbors of a given point in the dataset, and define:

$$\mu = \frac{r_2}{r_1} \geq 1$$

With some assumptions on the probability distribution of this sample, we can prove that this variable follows a *Pareto distribution*, that depends directly on the intrinsic dimension:

$$\Pr(\mu \leq t) = F_\mu(t) = \begin{cases} 0, & t < 1 \\ 1 - t^{-d}, & t \geq 1 \end{cases} \quad \text{Cumulative distribution function}$$

So we get the following relation between the cumulative distribution function (CDF) and the dimension:

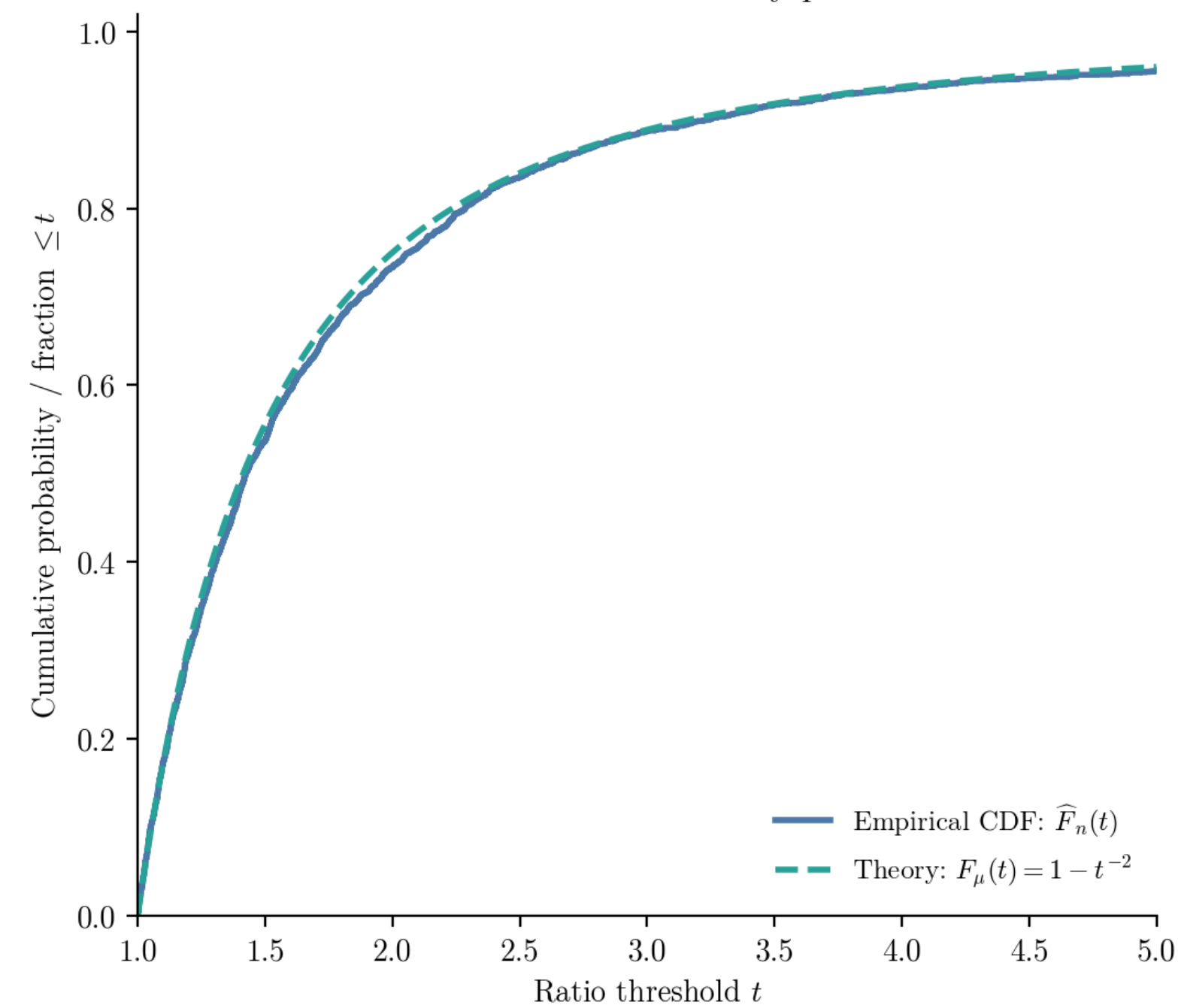
$$t^{-d} = 1 - F_\mu(t) \implies d \log t = -\log(1 - F_\mu(t))$$

Facco, E., d'Errico, M., Rodriguez, A., and Laio, A. Estimating the Intrinsic Dimension of Datasets by a Minimal Neighborhood Information. *Scientific Reports* 7, no. 1 (2017): 12140.

TwoNN

Approximating the dimension of the manifold

2. Pool ratios over many points



1. Collect the ratios for the points in your point cloud:

$$\mu_1, \dots, \mu_n, \quad \mu_i = \frac{r_{i,2}}{r_{i,1}}$$

2. Estimate the empirical CDF:

$$F_\mu(t) = \Pr(\mu \leq t) \qquad \hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\mu_i \leq t} \qquad \mathbf{1}_{\mu_i \leq t} = \begin{cases} 1, & \mu_i \leq t, \\ 0, & \text{otherwise.} \end{cases}$$

Initial point cloud: sample over the 2-dimensional sphere

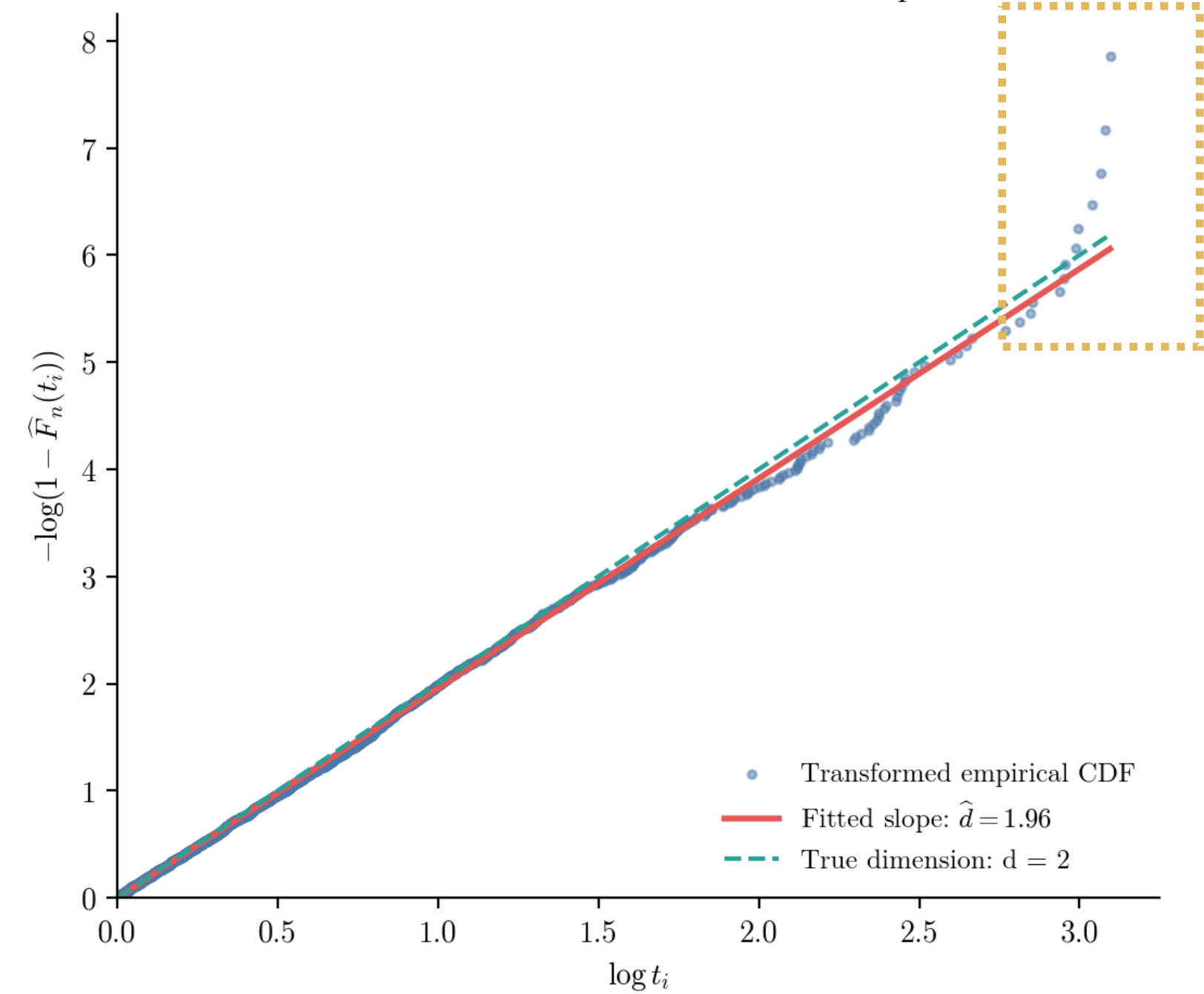
Facco, E., d’Errico, M., Rodriguez, A., and Laio, A. Estimating the Intrinsic Dimension of Datasets by a Minimal Neighborhood Information. *Scientific Reports* 7, no. 1 (2017): 12140.

TwoNN

Approximating the dimension of the manifold

Watch out for the extreme values that make the logarithm explode!

3. Estimate dimension from the slope



Initial point cloud: sample over the 2-dimensional sphere

1. Collect the ratios for the points in your point cloud: $\mu_1, \dots, \mu_n, \quad \mu_i = \frac{r_{i,2}}{r_{i,1}}$

2. Estimate the empirical CDF:

$$F_\mu(t) = \Pr(\mu \leq t) \qquad \hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\mu_i \leq t} \qquad \mathbf{1}_{\mu_i \leq t} = \begin{cases} 1, & \mu_i \leq t, \\ 0, & \text{otherwise.} \end{cases}$$

3. Order the ratios so that: $t_i = \mu_{(i)}$ (the smallest i -th ratio)

4. Fit a line to the points:

$$\left(\log t_i, \quad -\log\left(1 - \hat{F}_n(t_i)\right) \right)$$

The slope of the line should give you an estimate of the intrinsic dimension!

$$d \log t = -\log(1 - F_\mu(t))$$

Facco, E., d'Errico, M., Rodriguez, A., and Laio, A. Estimating the Intrinsic Dimension of Datasets by a Minimal Neighborhood Information. *Scientific Reports* 7, no. 1 (2017): 12140.

Maximum Likelihood Estimation (MLE)

Lecture 5 — Intrinsic dimension

MLE

Approximating the dimension of the manifold

What if we use the whole pattern of distances to the first k neighbors?

$$r_1 \leq r_2 \leq \dots \leq r_k = R \qquad u_i = \frac{r_i}{R}$$

Intuition: in higher dimensions, neighbors tend to be more concentrated near the boundary of the ball with the highest radius

We had seen that the probability of finding a point in a ball of radius r was $\Pr(X \in B^d(x, r)) \approx p(X)v_d r^d$ so

$$\Pr(U \leq t) = \frac{v_d(tR)^d}{v_d R^d} = t^d, \quad 0 \leq t \leq 1$$

*Probability of finding a point
with normalized radius*

$$U = \frac{r}{R} \leq t \iff r \leq tR$$

Example: the probability of finding a point in the inner half radius ball is 2^{-d} , so it decreases with dimension

Levina, E., and Bickel, P. (2004). Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17.

MLE

Approximating the dimension of the manifold

Maximum Likelihood Estimation: given a set of observations from a distribution with unknown parameters, what are the parameters that maximize the probability (likelihood) of having observed those observations?

Taking the derivative of the CDF before: $f_U(u \mid d) = dt^{d-1}$

Assuming all these observations are independent, the **likelihood** is:

$$L(d) \propto \prod_{j=1}^{k-1} du_j^{d-1}$$

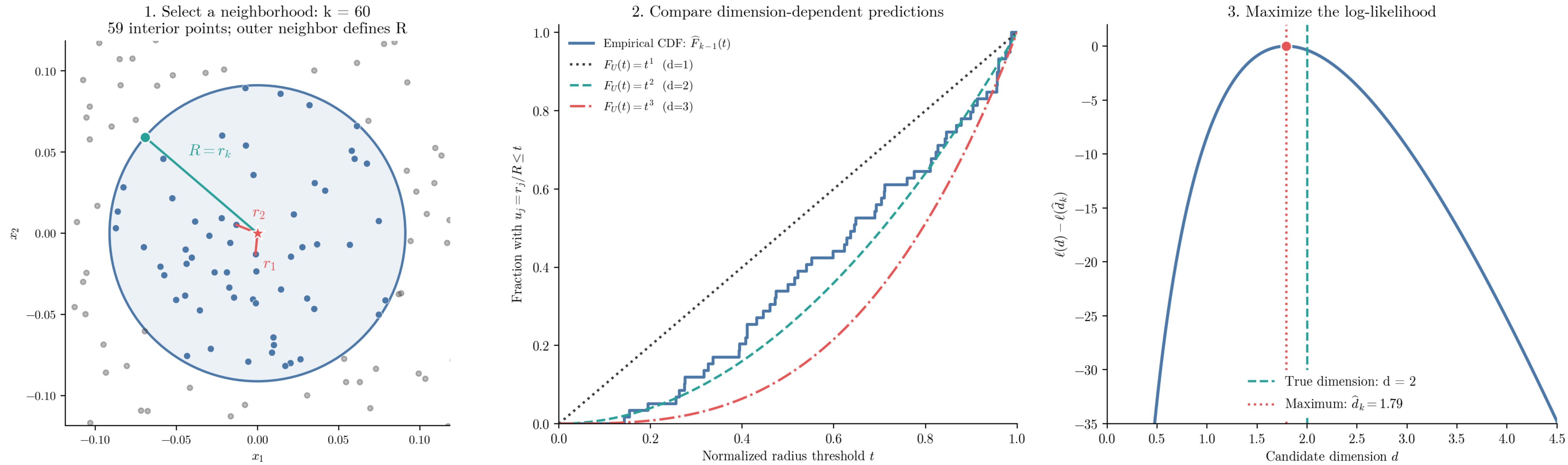
So the estimator that **maximizes the log-likelihood**: $\ell(d) = \log L(d) = (k-1) \log d + (d-1) \sum_{j=1}^{k-1} \log u_j + \text{constant}$

$$\hat{d}_k(x) = \frac{k-1}{\sum_{j=1}^{k-1} \log \left(\frac{r_k}{r_j} \right)}$$

Levina, E., and Bickel, P. (2004). Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17.

MLE

An example



Levina, E., and Bickel, P. (2004). Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17.

The manifold hypothesis in practice: Intrinsic dimension of image data

Lecture 5 — Intrinsic dimension

Intrinsic dimension of images

A case study

Let's go back to the initial example of the lecture:

What is the intrinsic dimension of image data? How does it affect deep learning methods?

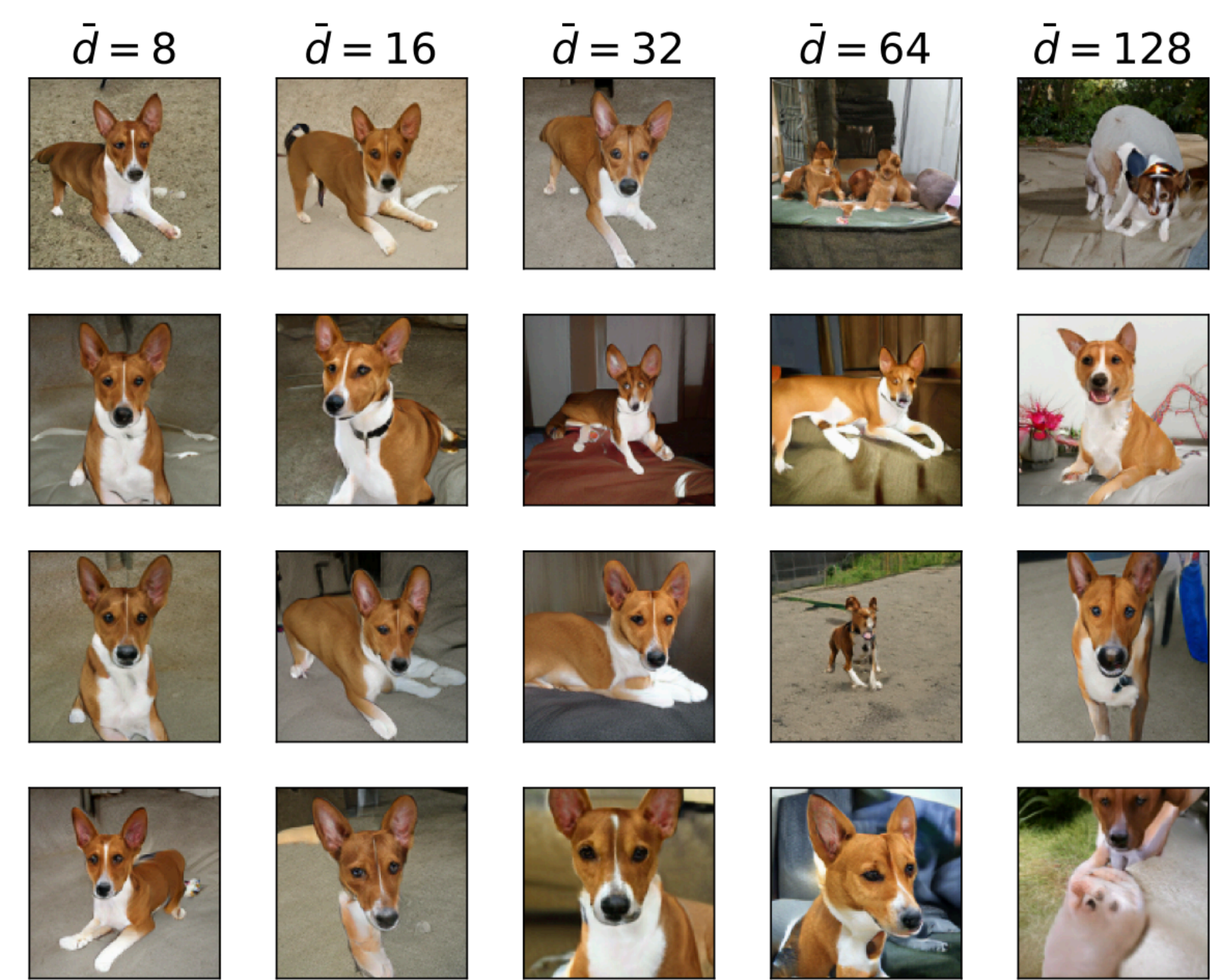
Main questions of the paper:

- Are MLE method accurate in estimating the intrinsic dimension of image data, where the ambient dimension is way higher than the intrinsic dimension?
- What are the estimated dimensions of common image datasets?
- Does the intrinsic dimension affect generalization properties? And the ambient dimension?

MLE methods for image data

Is the estimated dimension close to the expected dimension?

- Use a *Generative Adversarial Network* (GAN) trained on ImageNet to produce images from the class basenji
- In this model we can choose to fix most entries of the latent vectors to be 0, except for $\bar{d}$ entries which are chosen at random
- We expect that, as we increase $\bar{d}$, the intrinsic dimension of the generated data increases, keeping $\bar{d}$ as an upper bound



Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., and Goldstein, T. (2020). *The Intrinsic Dimension of Images and Its Impact on Learning*. International Conference on Learning Representations.

MLE methods for image data

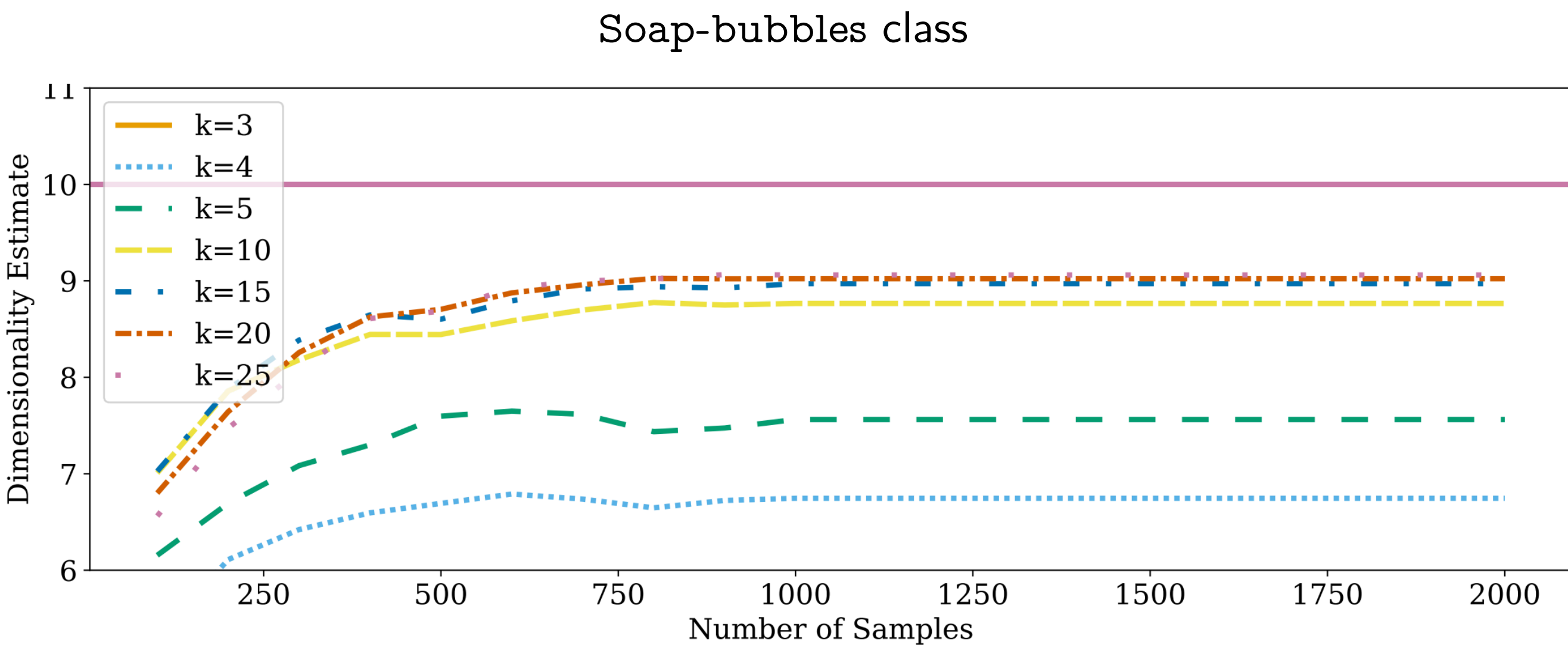
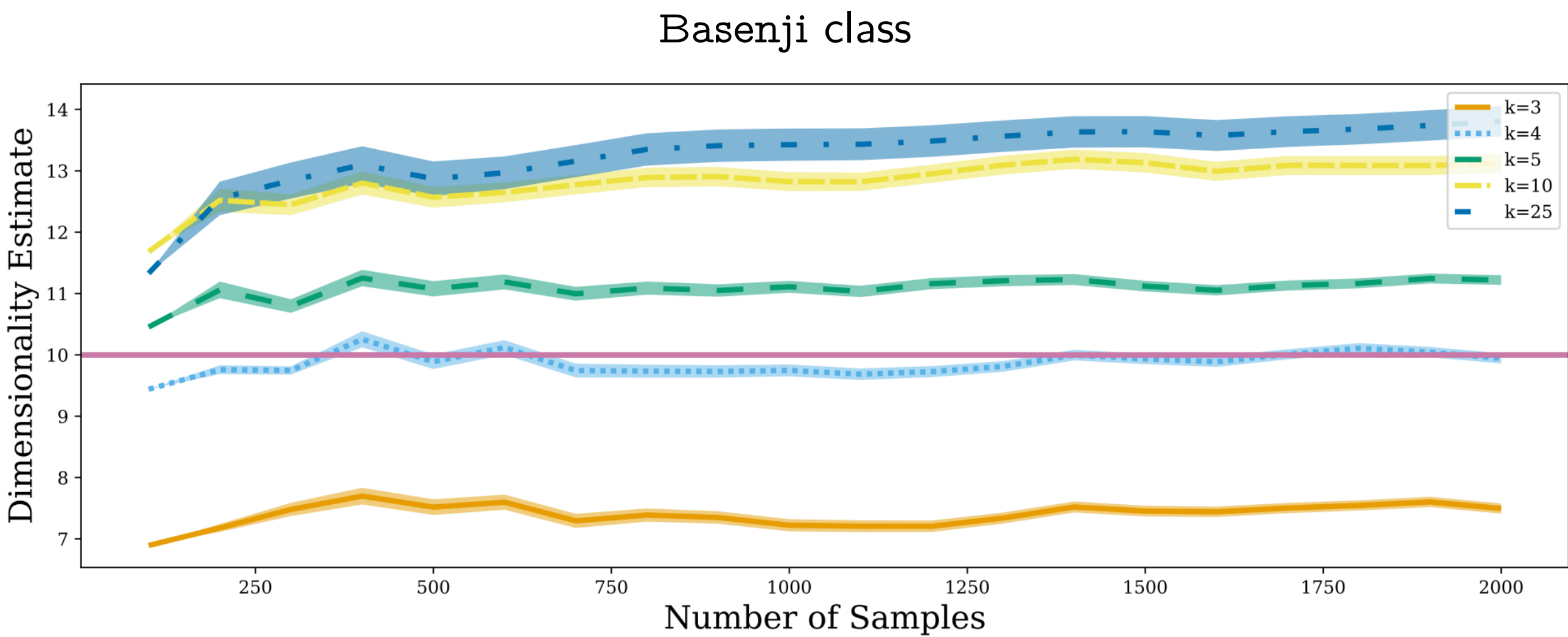
Is the estimated dimension close to the expected dimension?

Fix $\bar{d} = 10$, and generate an increasing number of examples for a given ImageNet class

Compute the intrinsic dimension using MLE with increasing number of neighbors (k)

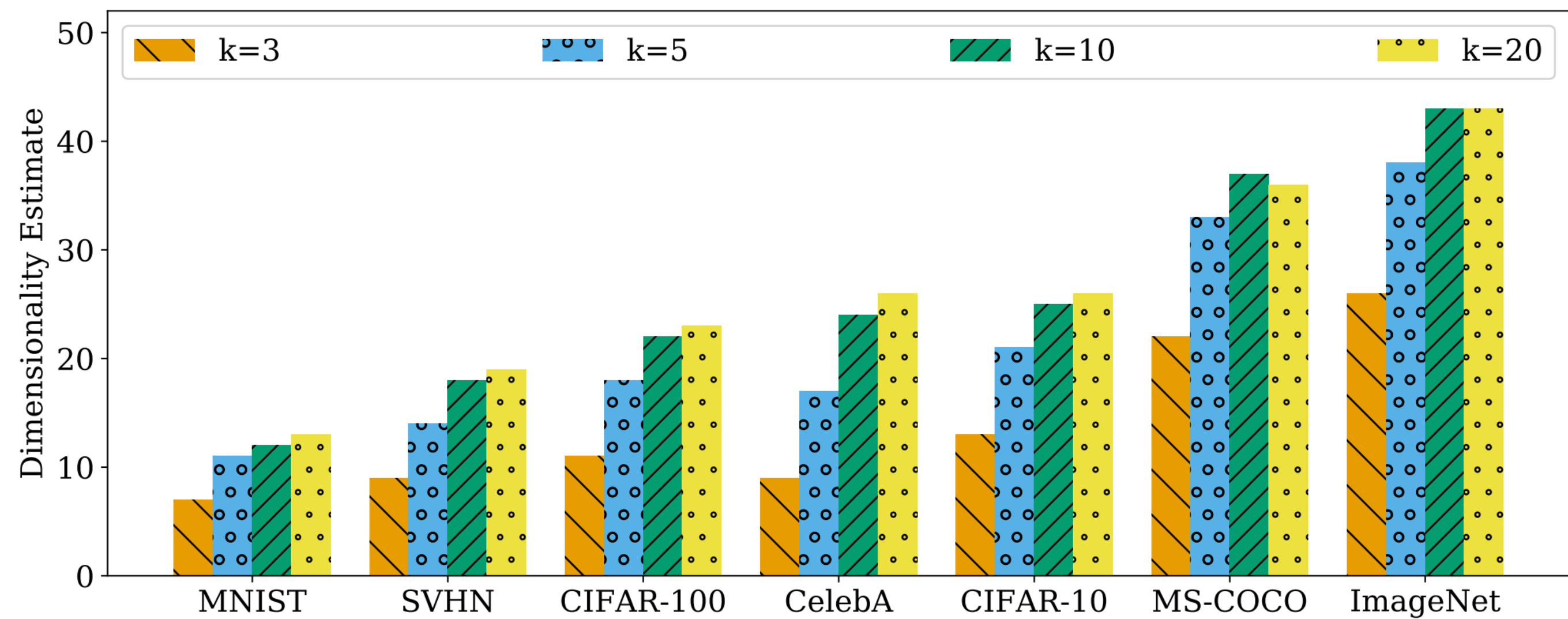
Note that, a priori, we do not know if these images are in the mathematical hypothesis of the estimators

As we increase the number of neighbors, the variance of the estimator decreases but the bias (difference between the average empirical estimated parameter and the true parameter) increases



Estimated dimension for popular datasets

What is the estimated dimension of popular image datasets?



Much smaller than the ambient dimension

Dataset / representation	Image dimensions	Ambient dimension
MNIST, standard grayscale	28 × 28 × 1	784
SVHN, CIFAR-10, CIFAR-100	32 × 32 × 3	3,072
ImageNet representation stated in the paper	224 × 224 × 3	150,528

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., and Goldstein, T. (2020). *The Intrinsic Dimension of Images and Its Impact on Learning*. International Conference on Learning Representations.

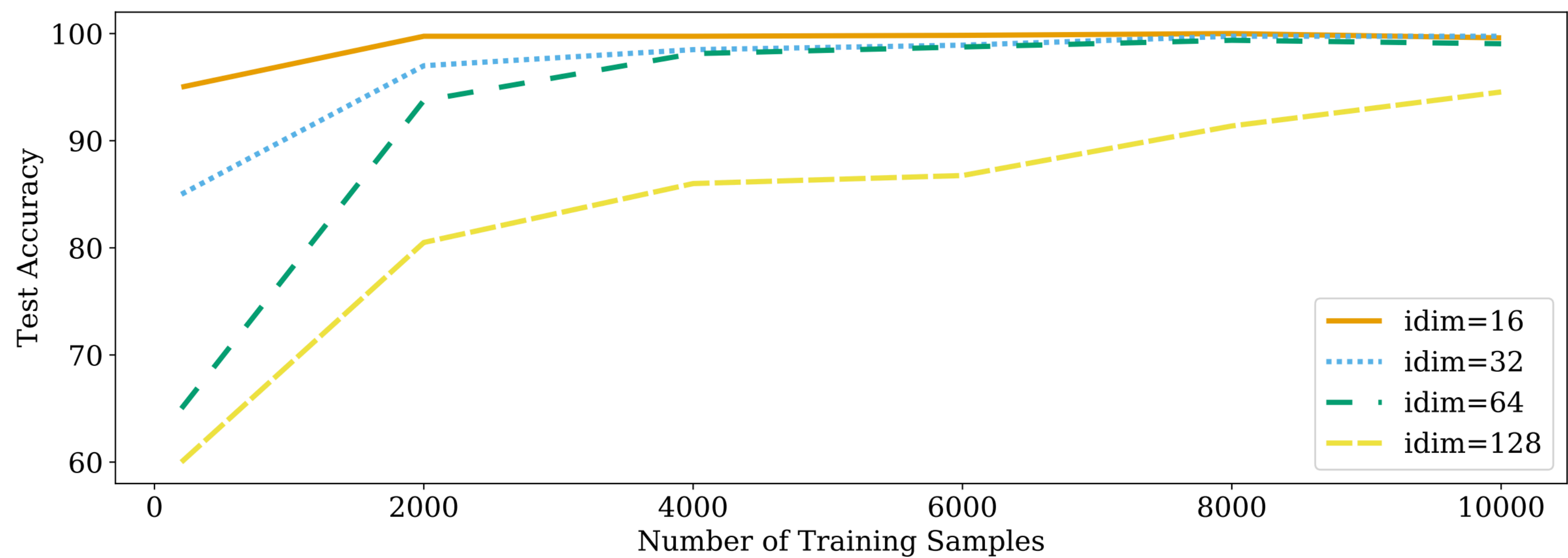
Intrinsic dimension and generalization

Does the intrinsic dimension of a dataset correlate with the generalization properties of a network trained on such data?

- Use the GAN to create 4 datasets of increasing intrinsic dimension: 16, 32, 64, 128
- These have fixed extrinsic dimension: all images have resolution $32 \times 32 \times 3$
- Create images from two classes: basenji and beagle
- Fix test size of 1,700 images, and train a ResNet-18 until 100% training accuracy, with increasing number of training samples
- Measure test accuracy

Intrinsic dimension and generalization

Does the intrinsic dimension of a dataset correlate with the generalization properties of a network trained on such data?



The lower the intrinsic dimension of the dataset that we are trying to learn, the easier it is for the network to learn the task

With higher dimensionalities, the network needs more training data to achieve good generalization properties

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., and Goldstein, T. (2020). *The Intrinsic Dimension of Images and Its Impact on Learning*. International Conference on Learning Representations.

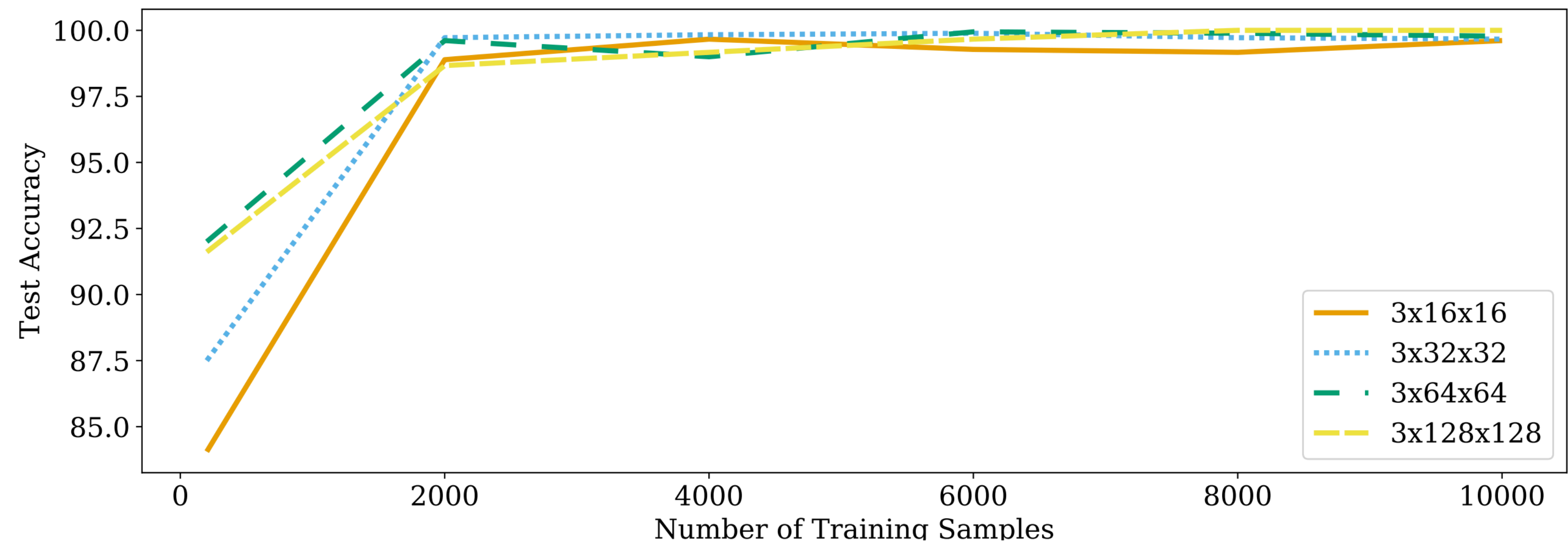
Extrinsic dimension and generalization

Does the extrinsic dimension of a dataset correlate with the generalization properties of a network trained on such data?

- Use nearest-neighbor interpolation to create 6 datasets of increasing extrinsic dimension: 3 channel images of resolution 16, 32, 64, 128, 256
- These have fixed intrinsic dimension of 128
- Create images from two classes: basenji and beagle
- Fix test size of 1,700 images, and train a ResNet-18 until 100% training accuracy, with increasing number of training samples
- Measure test accuracy

Extrinsic dimension and generalization

Does the extrinsic dimension of a dataset correlate with the generalization properties of a network trained on such data?



The lower the extrinsic dimension of the dataset does not have a strong influence on the generalization properties: all conditions arrive to near perfect generalization properties when we reach 2000 training samples

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., and Goldstein, T. (2020). *The Intrinsic Dimension of Images and Its Impact on Learning*. International Conference on Learning Representations.

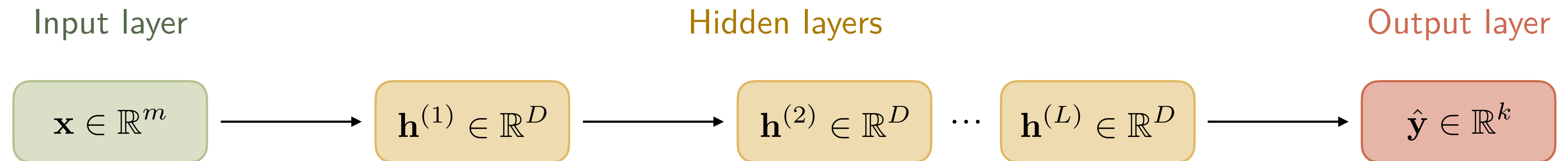
Intrinsic dimension of internal representations

Lecture 5 — Intrinsic dimension

Intrinsic dimension of internal representations in CNNs

How does the intrinsic dimension of latent representations behave?

$$f_{\theta} : \mathbb{R}^m \rightarrow \mathbb{R}^k$$



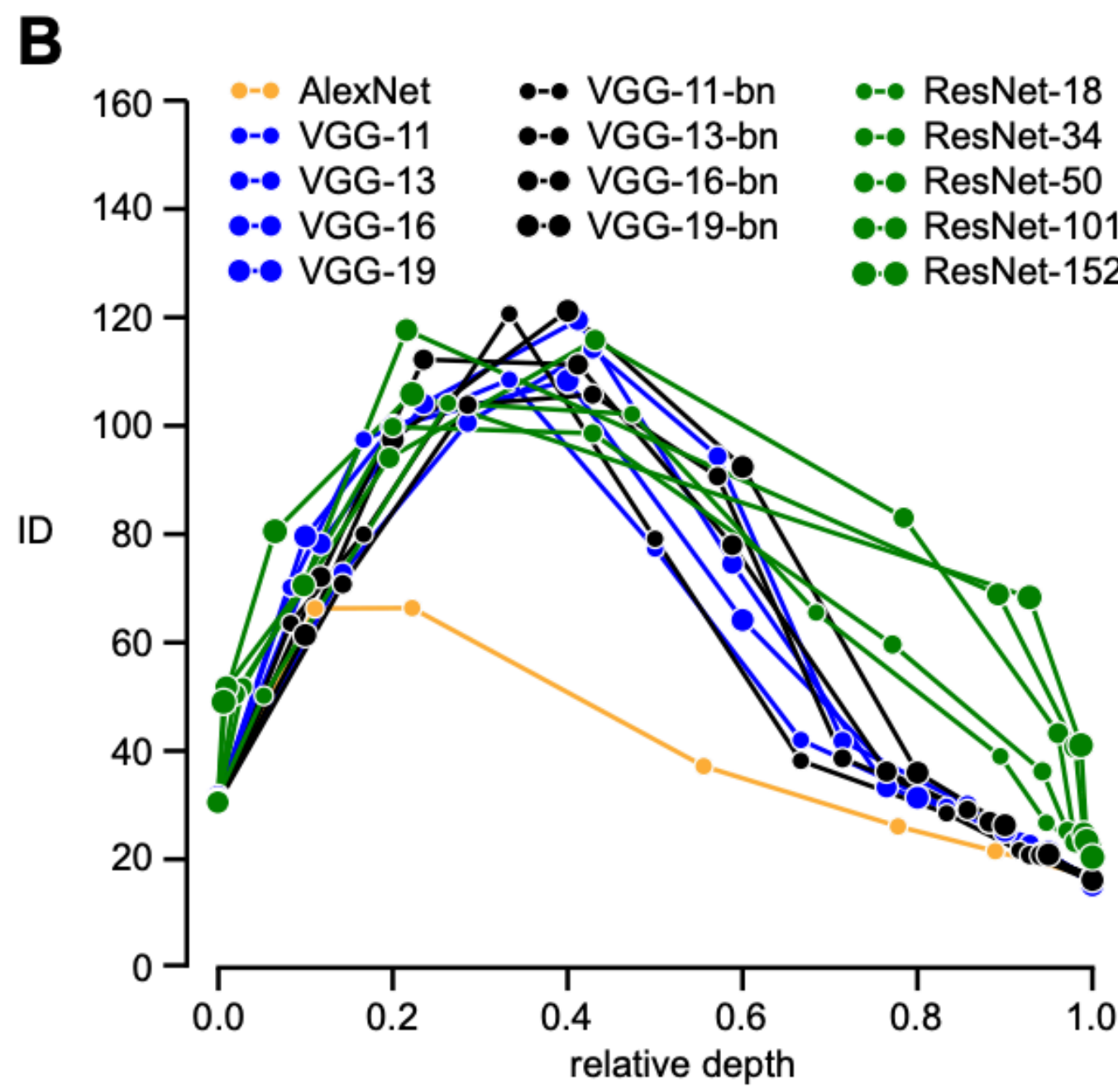
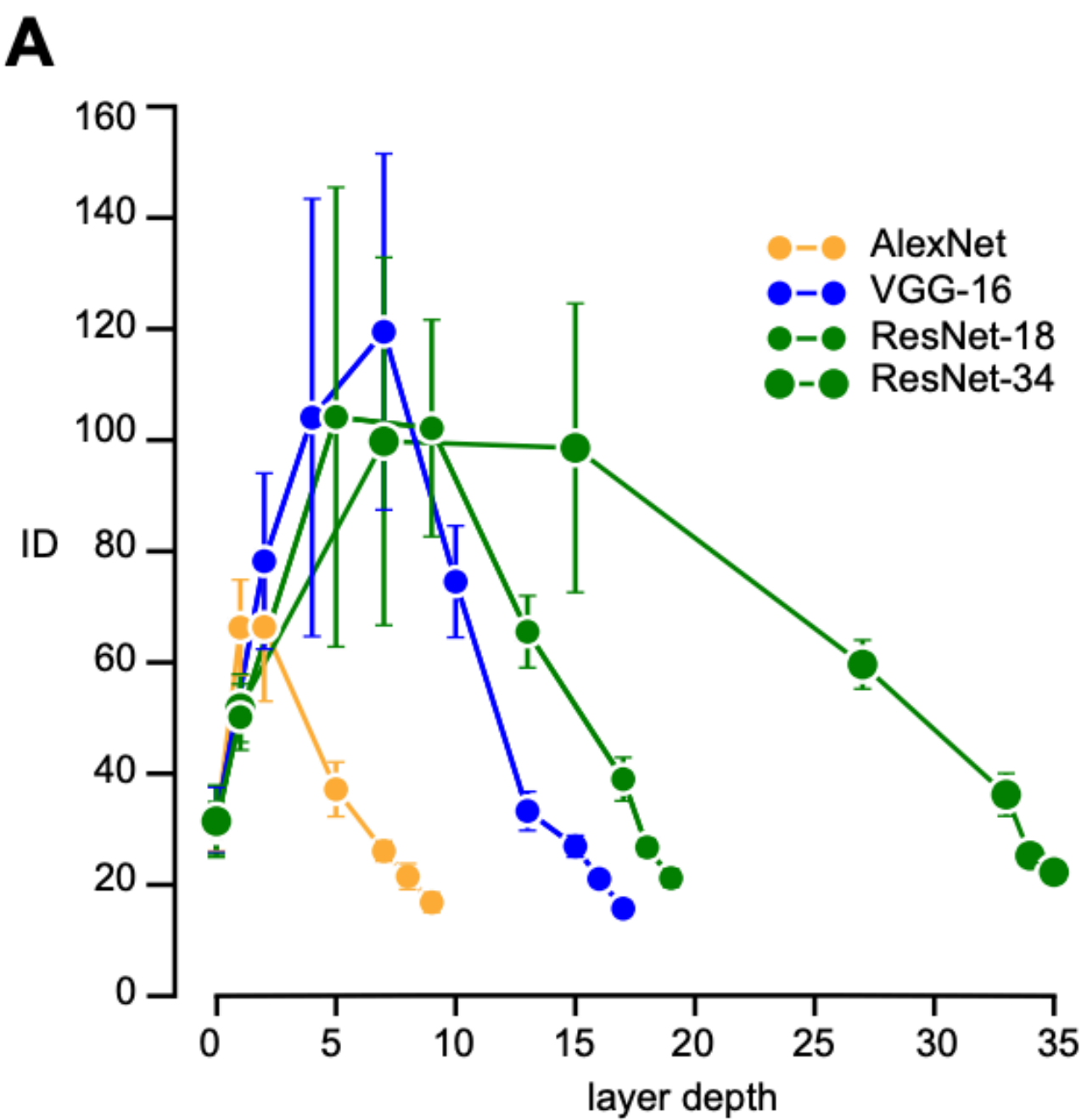
Main question: how does the intrinsic dimension of the internal representations change?

- Train different types of CNNs on ImageNet
- Take 500 images per category of the 7 biggest categories
- Compute embeddings at each layer: this gives 7 point clouds of 500 points after each layer
- Estimate the intrinsic dimension of the resulting point clouds using TwoNN
- Compute the average intrinsic dimension and standard deviations among classes, per layer

Ansuini, A., Laio, A., H. Macke, J., and Zoccolan, D. "Intrinsic Dimension of Data Representations in Deep Neural Networks." *Advances in Neural Information Processing Systems* 32 (2019).

The hunchback pattern

How does the intrinsic dimension of latent representations behave?



- Observed the appearance of a hunchback pattern: dimensionality goes up then it goes down
- In general, intrinsic dimension is way smaller than the hidden dimension of the models (ambient dimension in this case)
- The pattern is consistent across model families if we look at the relative depth:
$$\frac{\text{current layer}}{\text{total number of layers}}$$

In the last layer all models converge to a dimension of around 10

This is due to the fact that the representations in the last layer must be optimal for the classification task

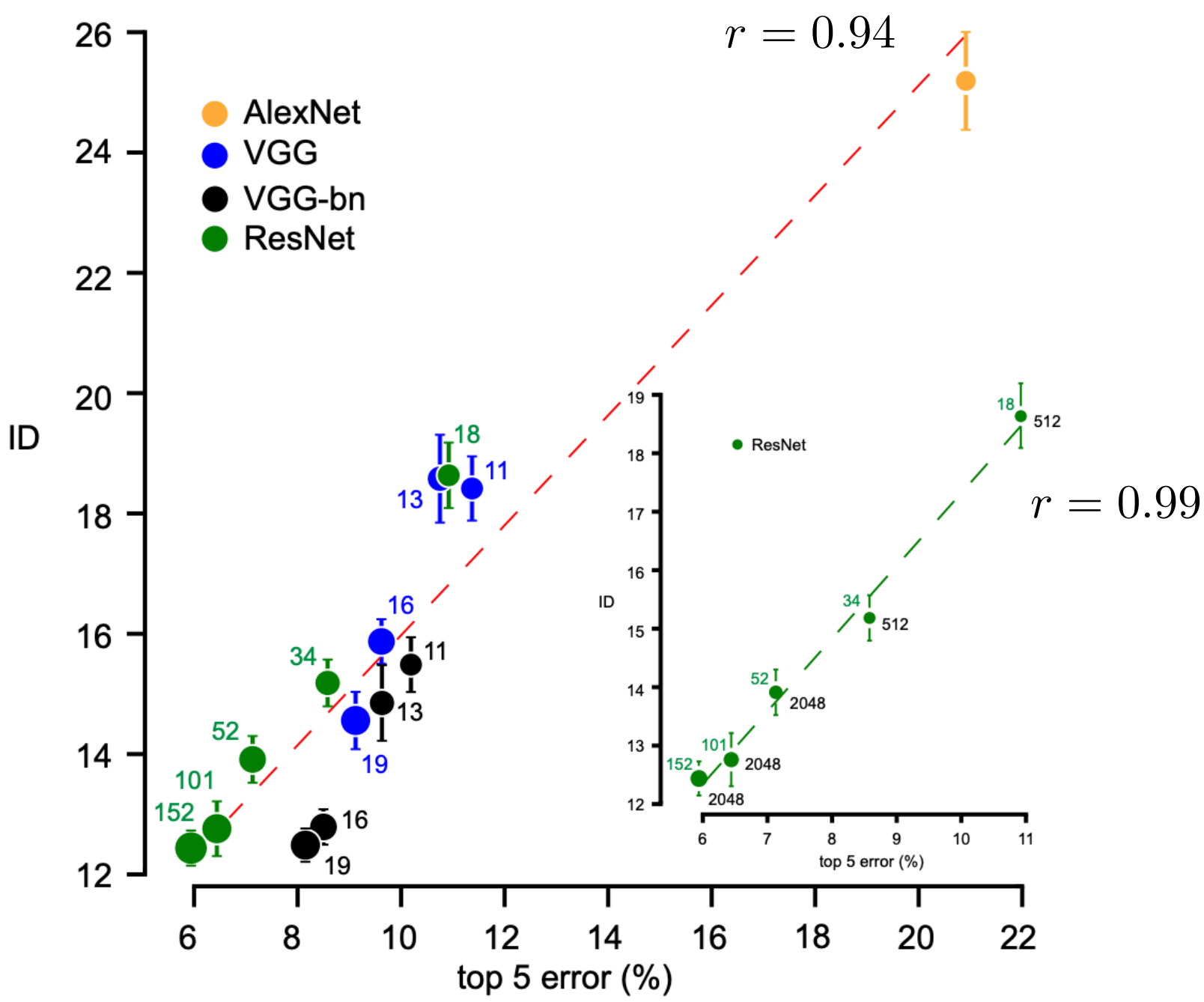
For N_c classes, the intrinsic dimension of a representation is bounded by $N_c \leq 2^{\text{dimension}}$, meaning that each class is associated with a binary representation

For the 1000 classes of ImageNet, this bound becomes dimension $\gtrsim 10$

Ansuini, A., Laio, A., H. Macke, J., and Zoccolan, D. "Intrinsic Dimension of Data Representations in Deep Neural Networks." *Advances in Neural Information Processing Systems* 32 (2019).

The intrinsic dimension in the last layer correlates with classification performance

How does the intrinsic dimension of latent representations behave?



- Compute the dimension of the representations in the last layer of 2,000 images, sampled from all classes in ImageNet
- Dimensions range from around 12 (ResNet152) to around 25 (AlexNet)
- Record the top-5-error of the networks (percentage of images whose correct class is absent from the model's five highest-scoring predictions)

The authors found strong correlation between the top-5-error and the intrinsic dimension: lower dimensionality means that the model is better at classifying the images

The correlation is particularly strong among networks of the same family

Ansuini, A., Laio, A., H. Macke, J., and Zoccolan, D. "Intrinsic Dimension of Data Representations in Deep Neural Networks." *Advances in Neural Information Processing Systems* 32 (2019).

Intrinsic dimension of internal representation in transformers

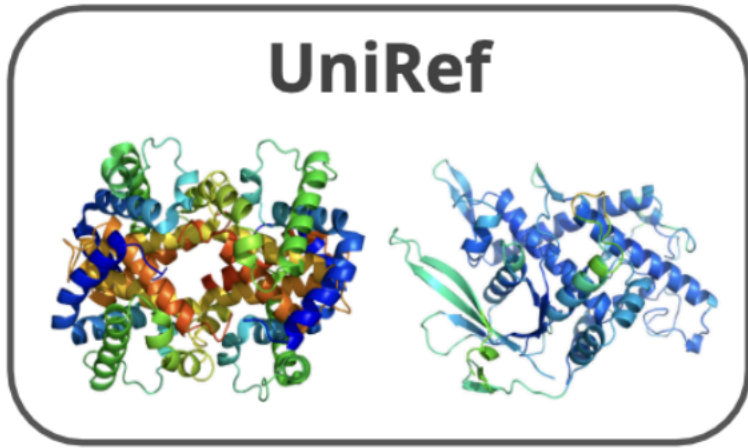
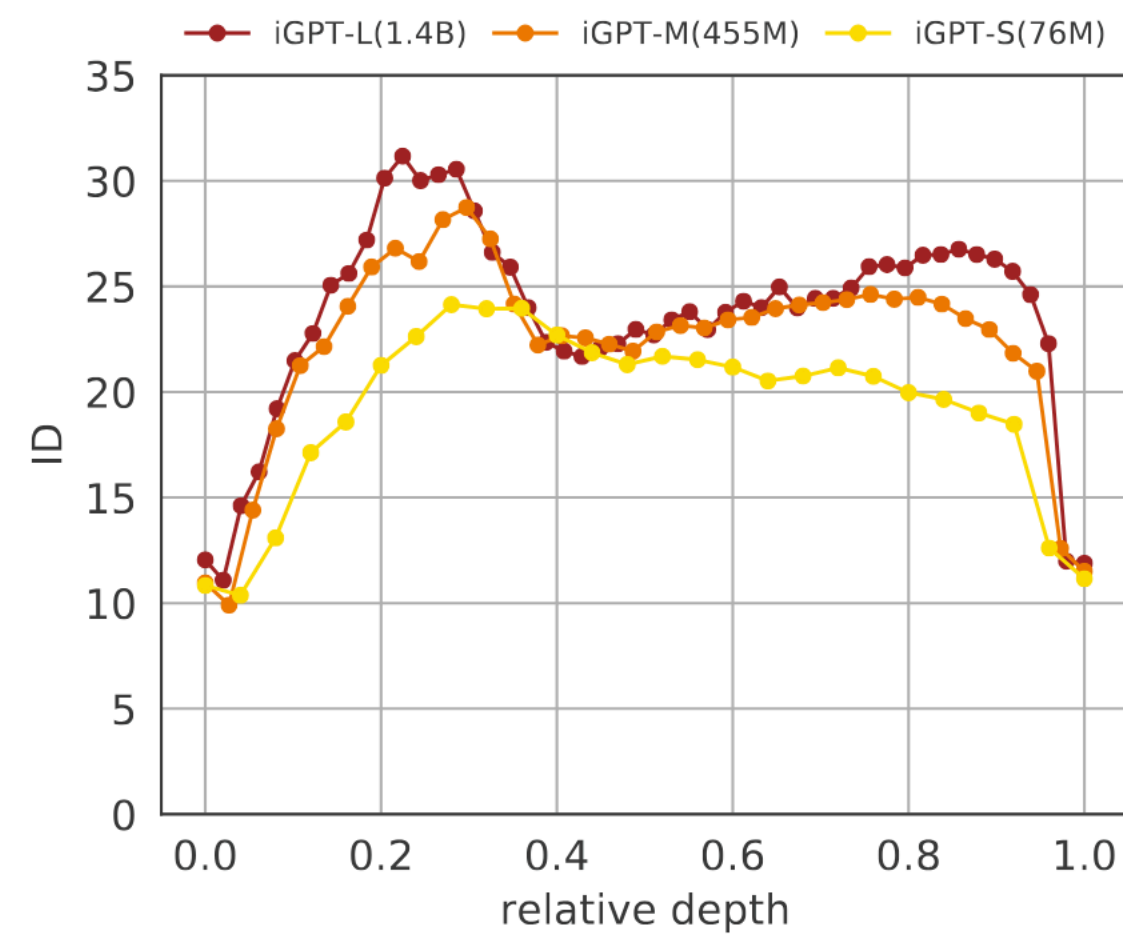
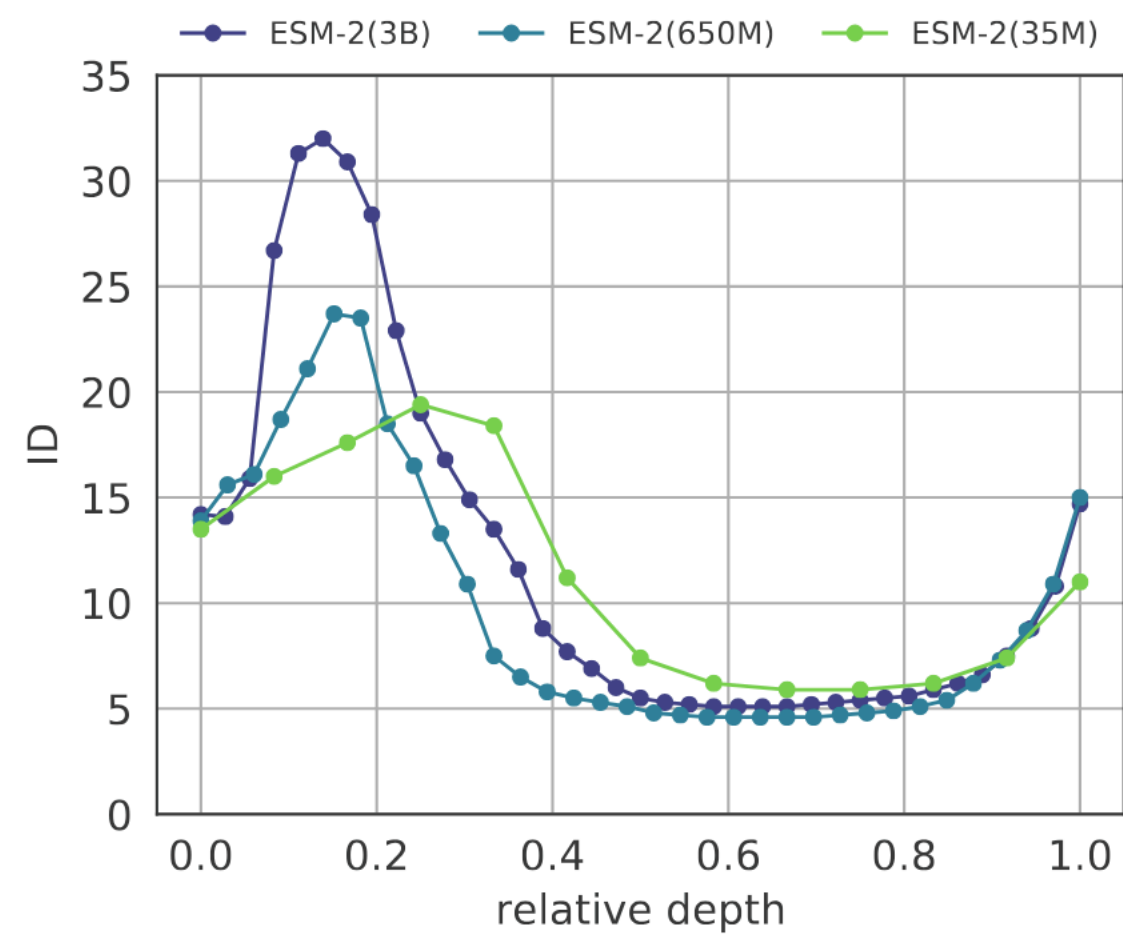
How does the intrinsic dimension of latent representations behave?

- Two transformer architectures: one for proteins, one for images
- Embed each dataset and add a positional encoding: obtain $\mathbb{R}^{T \times D}$ embeddings (one embedding per token)
- Vocabulary of 20 amino acids for proteins, and 512 colors for iGPTs
- Blocks: self-attention map followed by a MLP, output of size $\mathbb{R}^{T \times D}$
- Extract representations after the first normalization layer of each block and then average pool along the sequence dimension to reduce a sequence into a data point embedded in $\mathbb{R}^D$
- Compute intrinsic dimension using TwoNN estimator

Valeriani, L., Doimo, D., Cuturello, F., Laio, A., Ansuini, A., and Cazzaniga, A. “The Geometry of Hidden Representations of Large Transformer Models.” Paper presented at Thirty-seventh Conference on Neural Information Processing Systems. (2023)

The hunchback pattern reappears

How does the intrinsic dimension of latent representations behave?

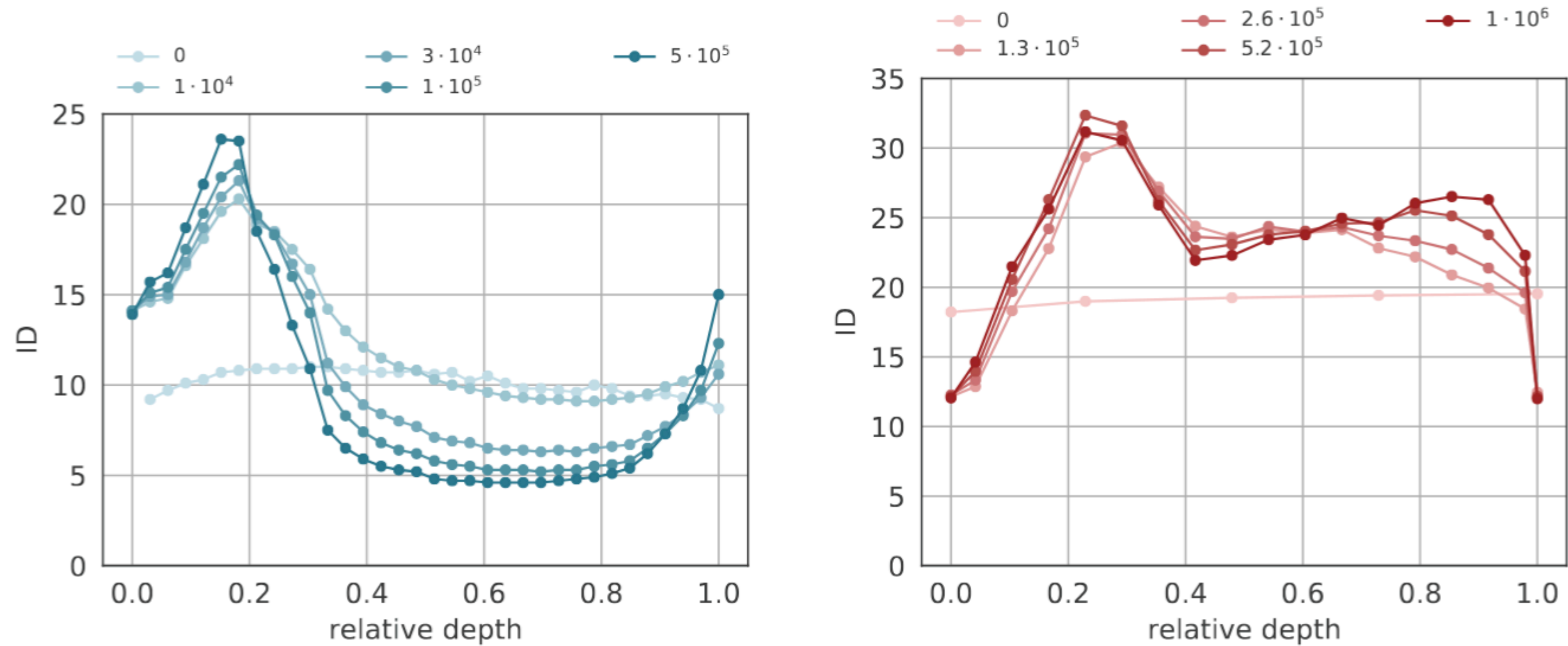


- For **proteins**: there is a peak that depends on the number of parameters of the model, then a plateau at around 5 for all models, then an increase to the initial ID of the model
A previous study of the dimensionality of this protein dataset had also estimated it around 5!
- For **images**: there is a second hunchback, and it also happens that the input and output dimensions coincide. The peak increases with model size, but all models coincide at the local minimum

Valeriani, L., Doimo, D., Cuturello, F., Laio, A., Ansuini, A., and Cazzaniga, A. "The Geometry of Hidden Representations of Large Transformer Models."
Paper presented at Thirty-seventh Conference on Neural Information Processing Systems. (2023)

Intrinsic dimension across training

How does the intrinsic dimension of latent representations behave?



It seems that there are **two stages** in the training:

1. First, early on, all models form a peak in the early layers, while the rest of the layer remain as in the untrained model
2. Later on, the plateau (or second hunchback) develops to its final value

Valeriani, L., Doimo, D., Cuturello, F., Laio, A., Ansuini, A., and Cazzaniga, A. "The Geometry of Hidden Representations of Large Transformer Models." Paper presented at Thirty-seventh Conference on Neural Information Processing Systems. (2023)

The union of manifolds hypothesis

Lecture 5 — Intrinsic dimension

The union of manifolds hypothesis

An extension of the manifold hypothesis

It might be a bit naive to suppose that all image lie on the same manifold, with the same intrinsic dimension

For example: if we understand the intrinsic dimension as the number of degrees of freedom in the data, this cannot be the same in the images of 2 and the images of 8 in the MNIST data

In addition, if we want to continuously transform an image of a 2 into an image of an 8 moving along the manifold, we might end in images that are neither are 2 nor an 8, leaving the dataset

Thus, it might be more accurate to describe the data as living in a union of manifolds, where each manifold has its own intrinsic dimension

This is known as the **union of manifolds hypothesis**

Brown, B. C., Caterini, A. L., Ross, B. L., Cresswell, J. C., and Loaiza-Ganem, G. (2022, September 29). “Verifying the Union of Manifolds Hypothesis for Image Data”. The Eleventh International Conference on Learning Representations.

The union of manifolds hypothesis in practice

An extension of the manifold hypothesis

To test whether the manifold hypothesis is satisfied in practice, we can start checking, for image data:

- Is the support of the data disconnected?
- Is there a lot of variability in the estimated dimensions (each connected component should have its own dimension)

Brown, B. C., Caterini, A. L., Ross, B. L., Cresswell, J. C., and Loaiza-Ganem, G. (2022, September 29). “Verifying the Union of Manifolds Hypothesis for Image Data”.
The Eleventh International Conference on Learning Representations.

Testing disconnectedness

The union of manifolds hypothesis

Table 1: FID scores. We show means and standard errors across 3 runs.

Model	MNIST	FMNIST	SVHN	CIFAR-10	CIFAR-100
WAE	19.1 ± 2.1	51.2 ± 2.4	89.8 ± 12.2	146.7 ± 1.2	145.0 ± 0.7
D-WAE (random)	23.7 ± 2.8	61.1 ± 3.6	77.6 ± 1.7	154.7 ± 0.5	146.5 ± 3.2
D-WAE (classes)	13.5 ± 0.2	36.8 ± 1.4	83.7 ± 15.2	133.3 ± 0.5	209.9 ± 3.5

The authors train a Wasserstein Autoencoder (WAE), which is unable to model datasets with disconnected supports, and introduce disconnected WAEs (D-WAE), as an alternative, where first cluster the data and then apply the model to each cluster

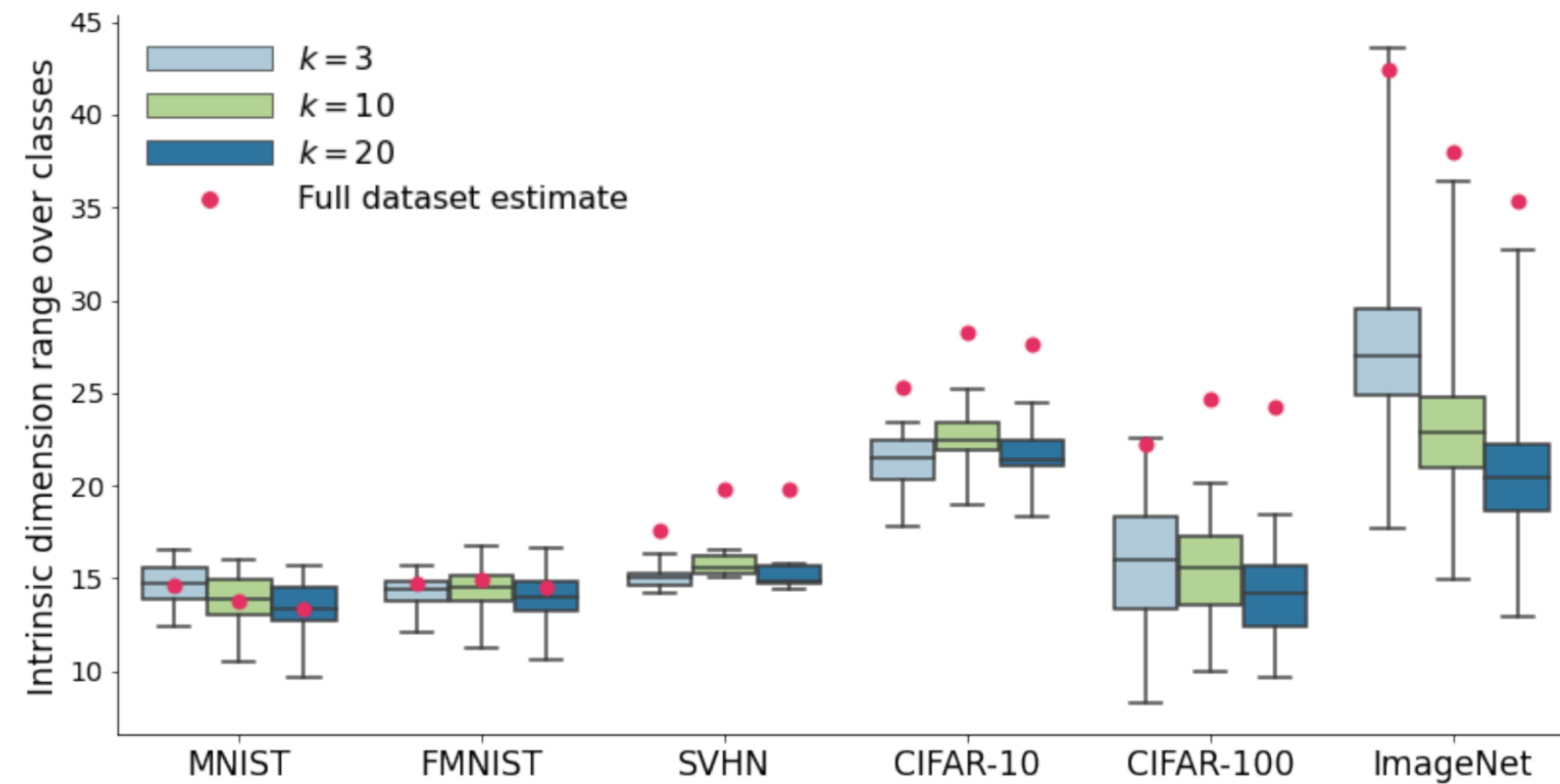
These are generative models, which they train in several popular image datasets

They report Fréchet Inception distance (FID, lower is better) across 3 runs, and observe that the best performing models are the disconnected ones

Brown, B. C., Caterini, A. L., Ross, B. L., Cresswell, J. C., and Loaiza-Ganem, G. (2022, September 29). “Verifying the Union of Manifolds Hypothesis for Image Data”. The Eleventh International Conference on Learning Representations.

Testing variability of dimension

The union of manifolds hypothesis



- Compute intrinsic dimension using MLE methods separately on the different point clouds given by the clusters of the classes $\hat{d}_k^c$
- Obtaining similar estimates across clusters supports the manifold hypothesis (constant dimension across the manifold), whereas a high variability in the estimates supports the union of manifolds hypothesis
- Except for SVHN, the box plots of this parameter support the latter

Brown, B. C., Caterini, A. L., Ross, B. L., Cresswell, J. C., and Loaiza-Ganem, G. (2022, September 29). “Verifying the Union of Manifolds Hypothesis for Image Data”. The Eleventh International Conference on Learning Representations.

Local Intrinsic Dimension in LLMs

Lecture 5 — Intrinsic dimension

Local vs. global measures of intrinsic dimension

A small recap

The estimators that we have seen here can be:

- **Global**, i.e., providing a single estimated dimension for the whole point cloud $\hat{d}$
- **Local**, i.e., computing an estimated dimension at each point by looking at the dimension at a local neighborhood, typically given by the k -nearest neighbors of the point $\hat{d}_k(x)$

Local	Global
?	?
?	

Local vs. global measures of intrinsic dimension

A small recap

The estimators that we have seen here can be:

- **Global**, i.e., providing a single estimated dimension for the whole point cloud $\hat{d}$
- **Local**, i.e., computing an estimated dimension at each point by looking at the dimension at a local neighborhood, typically given by the k -nearest neighbors of the point $\hat{d}_k(x)$

Local	Global
Local PCA	TwoNN
MLE estimators	

We can apply a similar idea to the TwoNN estimator to make it local:

- For each point x in the point cloud, take its k -nearest neighbors
- Compute the dimension of the neighborhood around x using the TwoNN estimator

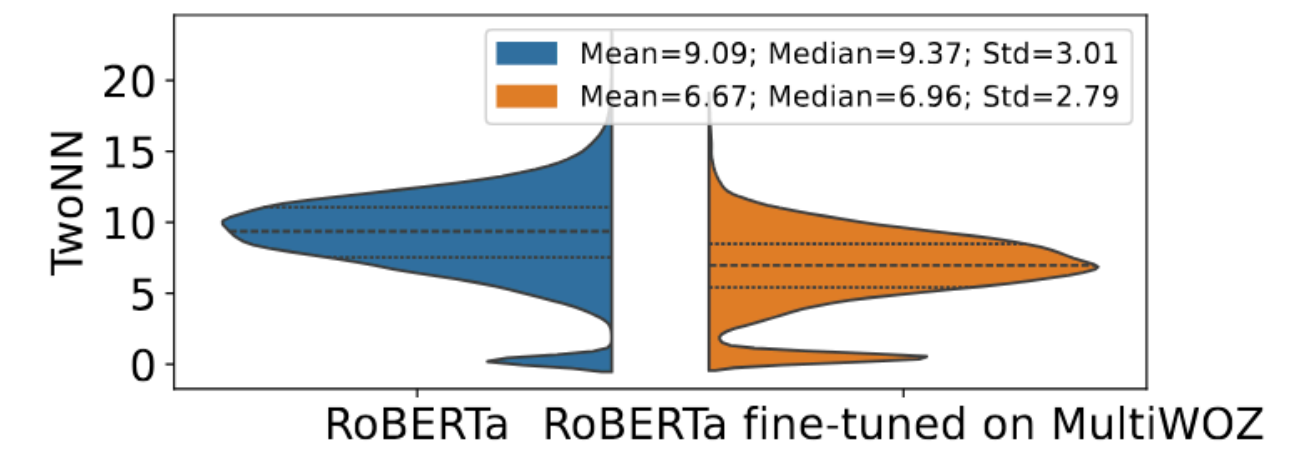
Ruppik, B. M., von Rohrscheidt, J., van Niekerk, C., Heck, M., Vukovic, R., Feng, S., Lin, H., Lubis, N., Rieck, B., Zibrowius, M., and Gasic, M. (2026). Less is More: Local Intrinsic Dimensions of Contextual Language Models. *Advances in Neural Information Processing Systems*, 38, 67795–67828.

Local TwoNN estimator of embeddings during LLM training

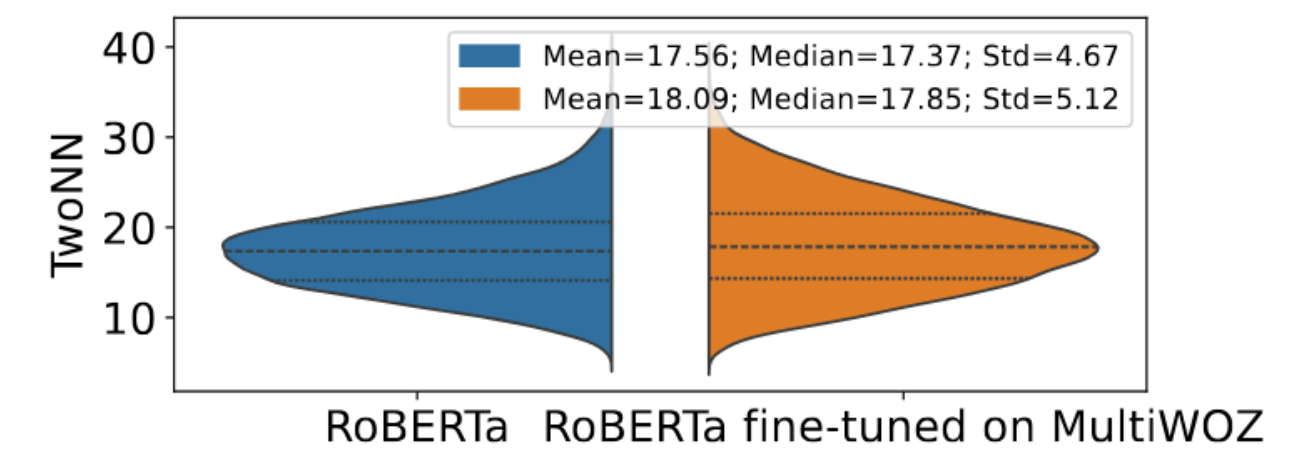
How does fine-tuning affect the local dimension of latent representations?

- Fine-tune a RoBERTa-base model using masked language modeling
- Each model is trained for 5 epochs on 10,000 training examples
- The base model is pre-trained on Wikipedia
- Fine-tuning is performed on a variety of datasets:
 - MultiWOZ2.1: Human-human multi-domain dialogues
 - Schema-Guided Dialogue SGD: Human–virtual assistant dialogues
- The authors embed a subset of the validation split of each dataset using both the base and fine-tuned model, and compute local TwoNN dimension estimates for each point
- For evaluation, the following datasets are also used:
 - Reddit: Reddit comments from the year 2022 mentioning Tesla, Inc.
 - ICLR 2024 Submissions: Titles and abstracts of ICLR 2024 papers

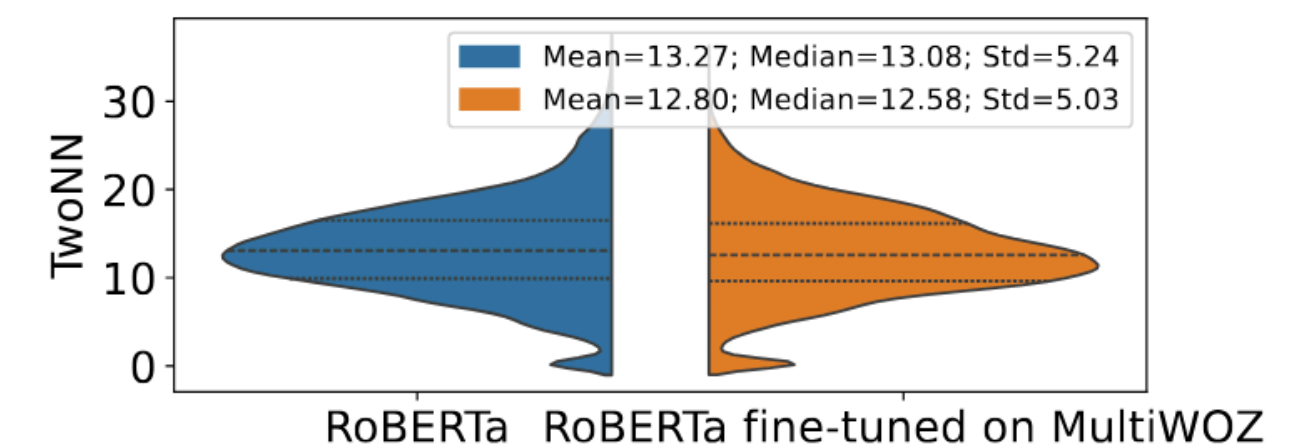
Main conclusion: only the intrinsic dimension of the datasets used in fine-tuning seems to change in this experiment



(a) TwoNN estimates on MultiWOZ validation



(b) TwoNN estimates on Wikipedia validation



(c) TwoNN estimates on Reddit validation

Ruppik, B. M., von Rohrscheidt, J., van Niekerk, C., Heck, M., Vukovic, R., Feng, S., Lin, H., Lubis, N., Rieck, B., Zibrowius, M., and Gasic, M. (2026). Less is More: Local Intrinsic Dimensions of Contextual Language Models. *Advances in Neural Information Processing Systems*, 38, 67795–67828.

To sum up

Today, we have learnt about:

- The manifold hypothesis, key to understand the power of current deep learning methods and the starting point of many geometric and topological methods in data
- How to measure the intrinsic dimension of point cloud data via 3 different methods:
 - Local PCA
 - TwoNN
 - MLE estimators
- How to apply this tools in deep learning to:
 - Explore the manifold hypothesis of image data and link it to network properties
 - Explore the dimensionality of internal representations in deep learning models
- Finally, we have seen how to extend this hypothesis to accommodate the possibility of varying dimension in the manifold, and we have covered an example that seems to support this in LLM training

Thank you for your attention!



AIDOS LAB
AI FOR DATA-ORIENTED SCIENCE

