

Token embeddings violate the manifold hypothesis

Michael Robinson (American University)

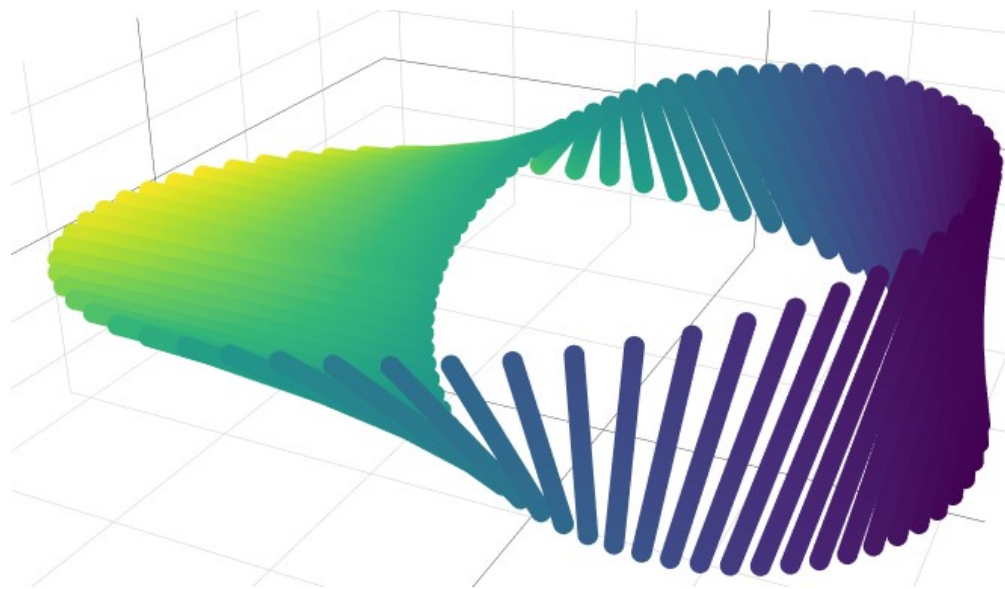
Sourya Dey (Galois, inc.)

Tony Chiang (University of Washington)

The manifold hypothesis

Data of interest are a sample drawn from the neighborhood of a manifold with low curvature

Consequence: There is a global dimension for each connected component



If the manifold hypothesis is true

- Then it's natural to try to find the manifolds → manifold learning!
- This seems implicit in LLM training, as LLMs explicitly use inner products
 - (Inner product spaces imply Riemannian manifolds)
- Assuming a manifold is similar to assuming a Gaussian distribution

What if the manifold hypothesis is not true?

- No notion of well-behaved geodesics
- Specific to LLMs, inner products will likely not preserve semantic similarity
 - Small perturbations in context may result in large variations in responses

Our contribution

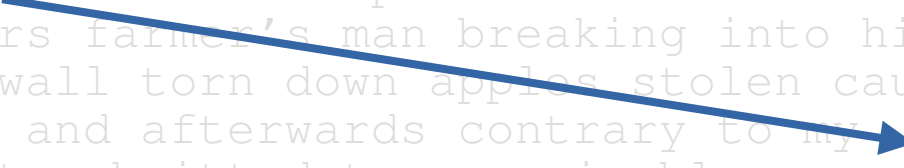
We can now test the manifold hypothesis

- 1) Two novel statistical tests (for manifolds and for fiber bundles)
- 2) Evidence to reject both hypotheses for several LLMs
- 3) The identification of otherwise unreported structure in the embeddings for these LLMs

no gentleman resident at monkford since the
time of old governor trent bless me how very
odd i shall forget my own name soon i suppose a
name that i am so very well acquainted with
knew the gentleman so well by sight seen him a
hundred times came to consult me once i
remember about a trespass of one of his
neighbours farmer's man breaking into his
orchard wall torn down apples stolen caught in
the fact and afterwards contrary to my
judgement submitted to an amicable compromise
very odd indeed after waiting another moment
you mean mr wentworth i suppose said anne mr
shepherd was all gratitude wentworth was the
very name mr wentworth was the very man he had
the curacy of monkford you know sir walter some
time back for two or three years came there
about the year

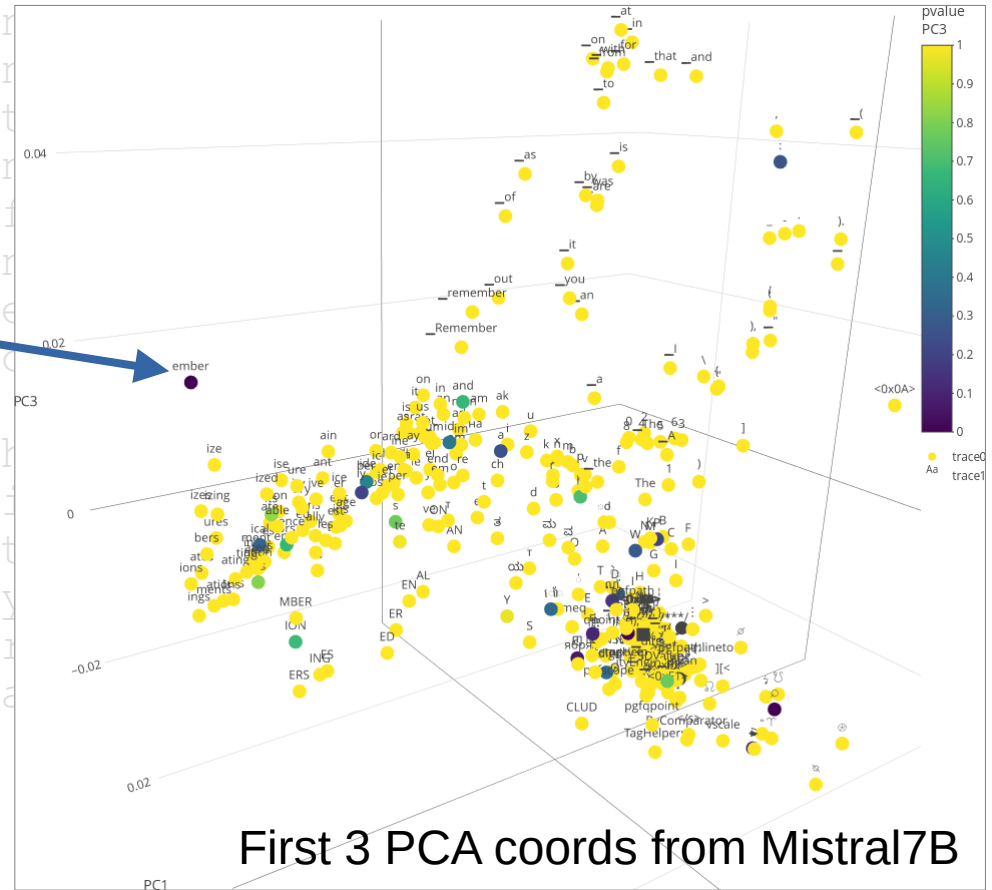
no gentleman resident at monkford since the
time of old governor trent bless me how very
odd i shall forget my own name soon i suppose a
name that i am so very well acquainted with
knew the gentleman so well by sight seen him a
hundred times came to consult me once i
remember about a trespass of one of his
neighbours farmer's man breaking into his
orchard wall torn down apples stolen caught in
the fact and afterwards contrary to my
judgement submitted to an amicable compromise
very odd indeed after waiting another moment
you mean mr wentworth i suppose said anne mr
shepherd was all gratitude wentworth was the
very name mr wentworth was the very man he had
the curacy of monkford you know sir walter some
time back for two or three years came there
about the year

no gentleman resident at monkford since the
time of old governor trent bless me how very
odd i shall forget my own name soon i suppose a
name that i am so very well acquainted with
knew the gentleman so well by sight seen him a
hundred times came to consult me once i
remember about a trespass of one of his
neighbours farmer's man breaking into his
orchard wall torn down apples stolen caught in
the fact and afterwards contrary to my
judgement submitted to an amicable compromise
very odd indeed after waiting another moment
you mean mr wentworth i suppose said anne mr
shepherd was all gratitude wentworth was the
very name mr wentworth was the very man he had
the curacy of monkford you know sir walter some
time back for two or three years came there
about the year



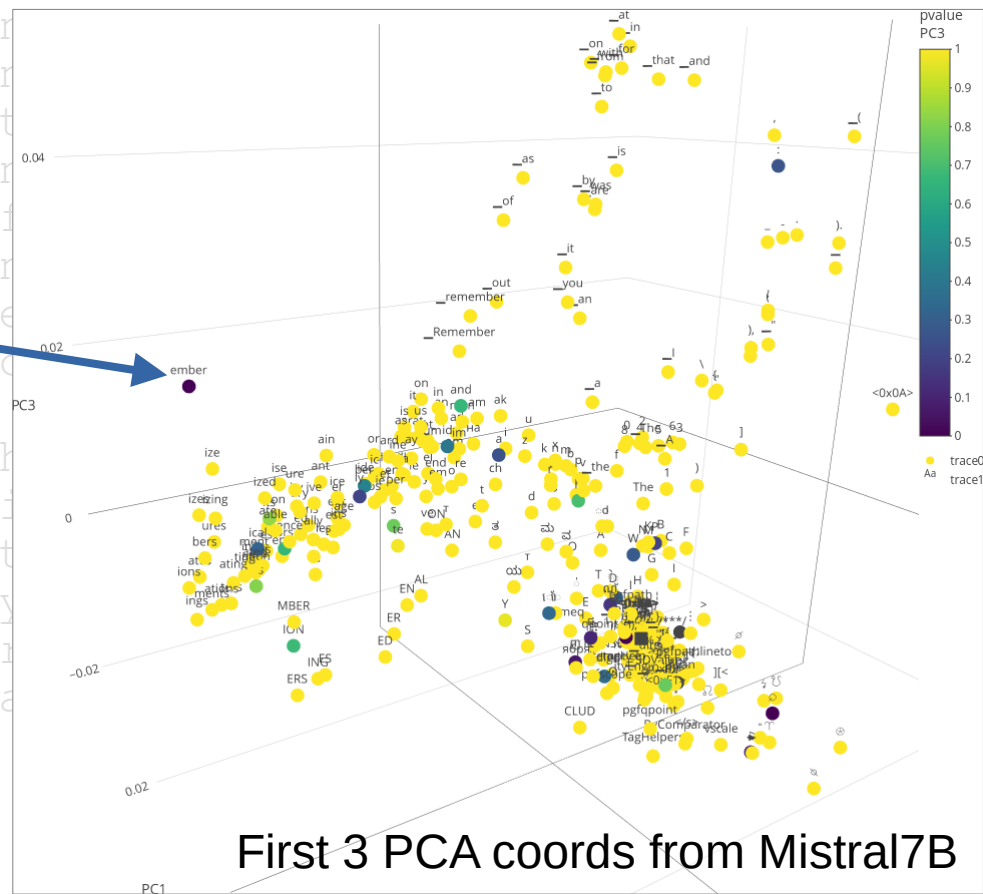
[-0.0018, -0.0033, 0.0014, ...]

no gentleman resident at monkford since the
time of old governor trent bless me how very
odd i shall forget my own name soon
name that i am so very well acquainted
knew the gentleman so well by sight
hundred times came to consult me or
remember about a trespass of one of
neighbours farmer's man breaking in
orchard wall torn down apples stole
the fact and afterwards contrary to
judgement submitted to an amicable
very odd indeed after waiting another
you mean mr wentworth i suppose said
shepherd was all gratitude wentworth
very name mr wentworth was the very
the curacy of monkford you know since
time back for two or three years came
about the year



Is the set of tokens a sample from a low curvature, low dimensional manifold?

no gentleman resident at monkford since the time of old governor trent bless me how very odd i shall forget my own name soon name that i am so very well acquairt i knew the gentleman so well by sight hundred times came to consult me or remember about a trespass of one of neighbours farmer's man breaking in orchard wall torn down apples stole the fact and afterwards contrary to judgement submitted to an amicable very odd indeed after waiting another you mean mr wentworth i suppose said shepherd was all gratitude wentworth very name mr wentworth was the very the curacy of monkford you know six time back for two or three years came about the year



First 3 PCA coords from Mistral7B

Our statistical tests

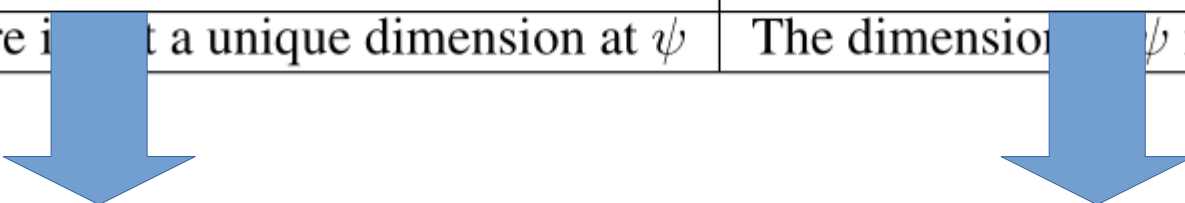
For each point (token) ψ we posit two pairs of classical hypothesis tests:

	Manifold test	Fiber bundle test
H_0	There is a unique dimension at ψ	The dimension at ψ in a ball of radius r does not increase as r increases.
H_1	There is not a unique dimension at ψ	The dimension at ψ increases at some r

Our statistical tests

For each point (token) ψ we posit two pairs of classical hypothesis tests:

	Manifold test	Fiber bundle test
H_0	There is a unique dimension at ψ	The dimension at ψ in a ball of radius r does not increase as r increases.
H_1	There is not a unique dimension at ψ	The dimension at ψ increases at some r

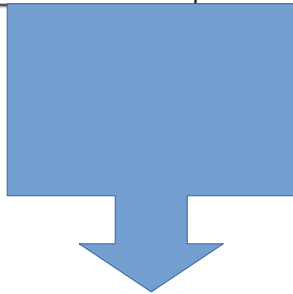


Small change in prompt results in small change in response

Our statistical tests

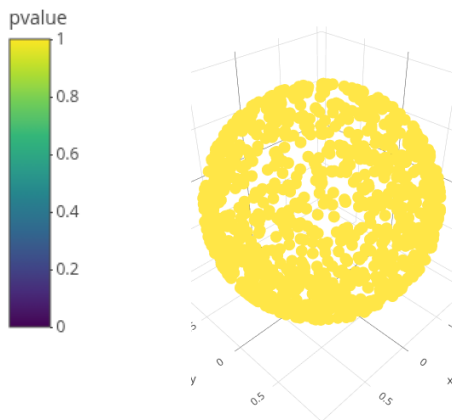
For each point (token) ψ we posit two pairs of classical hypothesis tests:

	Manifold test	Fiber bundle test
H_0	There is a unique dimension at ψ	The dimension at ψ in a ball of radius r does not increase as r increases.
H_1	There is not a unique dimension at ψ	The dimension at ψ increases at some r



Small changes in prompt may result in large changes in response

Testing for manifolds

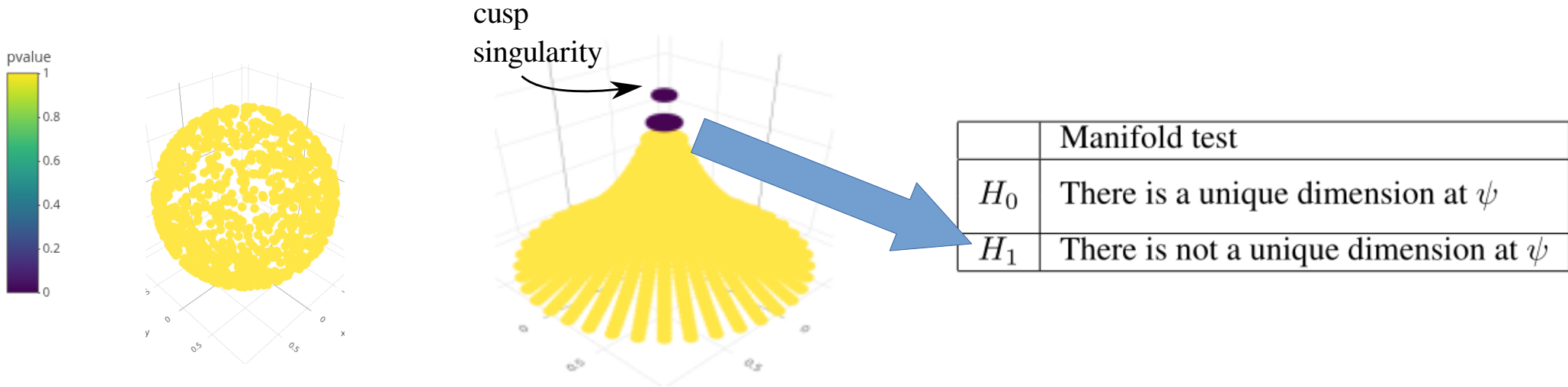


	Manifold test
H_0	There is a unique dimension at ψ
H_1	There is not a unique dimension at ψ

A sphere is a manifold, with a constant dimension...

Our test has high p -value for H_0 at every point,
so we **fail to reject** the manifold hypothesis

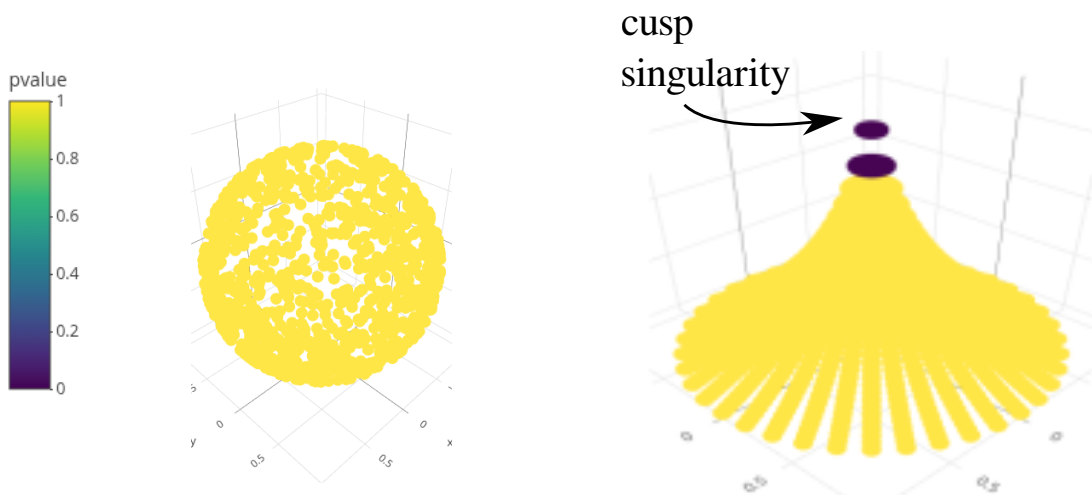
Testing for manifolds



A surface with a cusp singularity does not have a well defined dimension at the cusp... **reject** the manifold hypothesis H_0

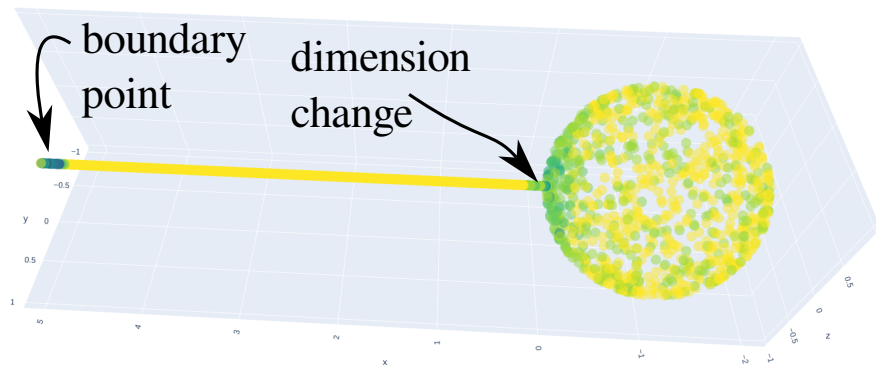
Even one rejection means the **entire space** is not a manifold!

Testing for manifolds

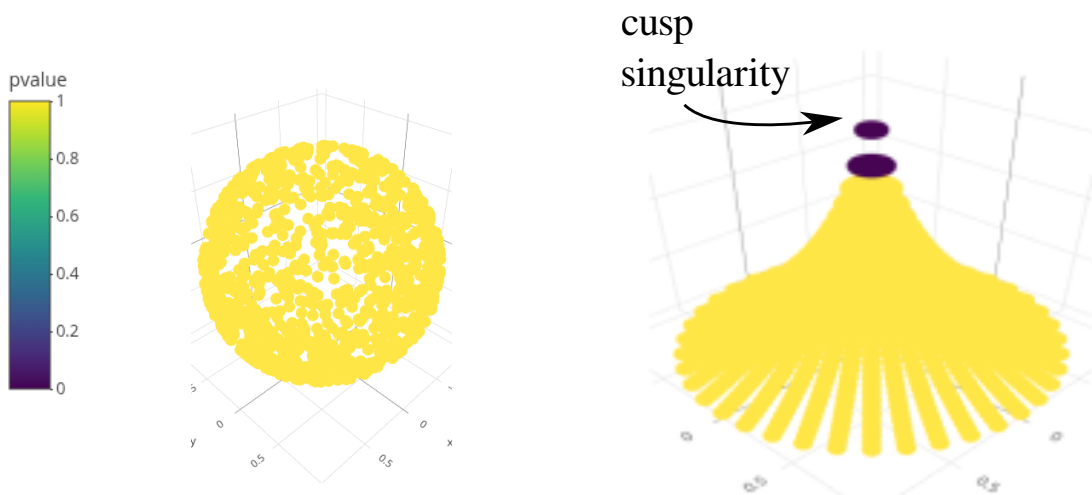


	Manifold test
H_0	There is a unique dimension at ψ
H_1	There is not a unique dimension at ψ

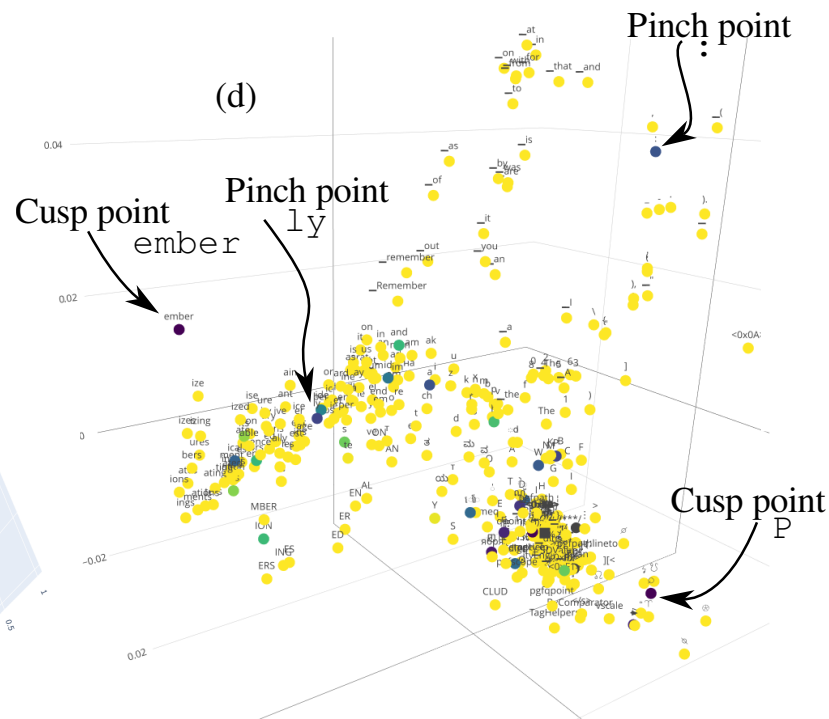
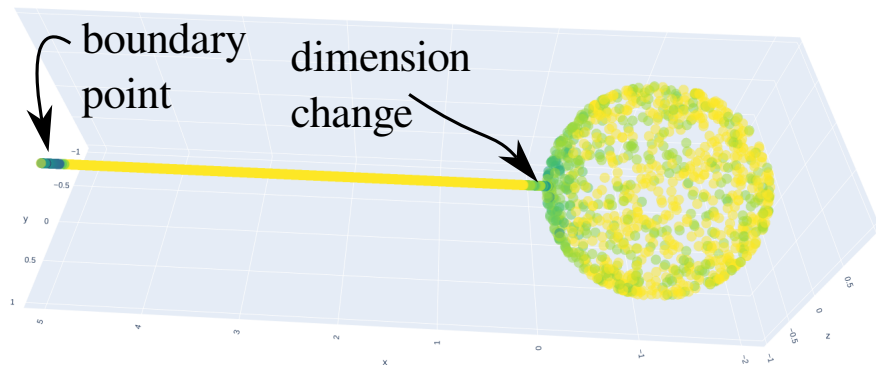
A lollipop space has two places where the manifold hypothesis is rejected, and for two different reasons.



Testing for manifolds



LLM token spaces exhibit many rejections of the manifold hypothesis!



First 3 PCA coords from Mistral7B

Model	Manifold rejects	Fiber bundle			
		Smaller Radius dim.	rejects	Larger radius dim.	rejects
GPT2 $n = 50257$	66 $p \approx 3 \times 10^{-8}$	Q1: 20 Q2: 389 Q3: 531	12 $p \approx 9 \times 10^{-6}$	Q1: 8 Q2: 14 Q3: 32	7 $p \approx 3 \times 10^{-8}$
Llemma7B $n = 32016$	33 $p \approx 5 \times 10^{-9}$	Q1: 4096 Q2: 4096 Q3: 4096	0 N/A	Q1: 8 Q2: 11 Q3: 14	1 $p \approx 3 \times 10^{-4}$
Mistral7B $n = 32016$	40 $p \approx 3 \times 10^{-7}$	Q1: 9 Q2: 48 Q3: 220	1 $p \approx 8 \times 10^{-4}$	Q1: 5 Q2: 6 Q3: 9	2 $p \approx 8 \times 10^{-5}$
Pythia6.9B $n = 50254$	54 $p \approx 2 \times 10^{-7}$	Q1: 2 Q2: 108 Q3: 235	0 N/A	Q1: 2 Q2: 5 Q3: 145	0 N/A

Model	Manifold rejects	Fiber bundle			
		Smaller Radius dim.	rejects	Larger radius dim.	rejects
GPT2 $n = 50257$	$\overset{66}{p \approx 3 \times 10^{-8}}$	Q1: 20 Q2: 389 Q3: 531	$\overset{12}{p \approx 9 \times 10^{-6}}$	Q1: 8 Q2: 14 Q3: 32	$\overset{7}{p \approx 3 \times 10^{-8}}$
Llemma7B $n = 32016$	$\overset{33}{p \approx 5 \times 10^{-9}}$	Q1: 4096 Q2: 4096 Q3: 4096	0 N/A	Q1: 8 Q2: 11 Q3: 14	$\overset{1}{p \approx 3 \times 10^{-4}}$
Mistral7B $n = 32016$	$\overset{40}{p \approx 3 \times 10^{-7}}$	Q1: 9 Q2: 48 Q3: 220	$\overset{1}{p \approx 8 \times 10^{-4}}$	Q1: 5 Q2: 6 Q3: 9	$\overset{2}{p \approx 8 \times 10^{-5}}$
Pythia6.9B $n = 50254$	$\overset{54}{p \approx 2 \times 10^{-7}}$	Q1: 2 Q2: 108 Q3: 235	0 N/A	Q1: 2 Q2: 5 Q3: 145	0 N/A

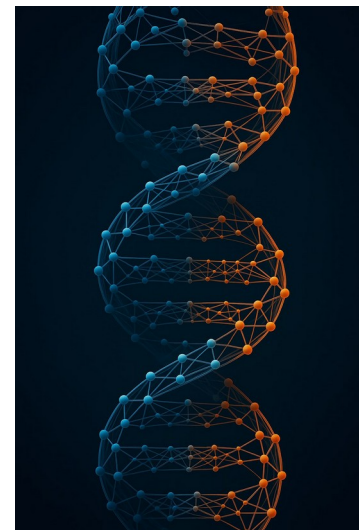
Implications

- Token embeddings are **not** samples from a low curvature manifold nor from a fiber bundle
- Irregular tokens may result in instabilities in LLM output
 - Small changes in a prompt may result in large changes in the response
 - Longer context windows and fine tuning do not resolve these instabilities (Thm 2)
- This may explain LLM behavioral features:
 - Glitch tokens
 - Behavior differences between models
- Much work remains to explore additional implications of our findings!
 - For instance, check out Vargas *et al.*, “Understanding Generative AI Content with Embedding Models,” *Science Advances*, Nov 2025.

Code at: https://github.com/kb1dds/stratified_estimator/

Implications

- Token embeddings are **not** samples from a low curvature manifold nor from a fiber bundle
- Irregular tokens may result in instabilities in LLM output
 - Small changes in a prompt may result in large changes in the response
 - Longer context windows and fine tuning do not resolve these instabilities (Thm 2)
- This may explain LLM behavioral features:
 - Glitch tokens
 - Behavior differences between models
- Much work remains to explore additional implications of our findings!
 - For instance, check out **Vargas *et al.*, “Understanding Generative AI Content with Embedding Models,” *Science Advances*, Nov 2025.**



Code at: https://github.com/kb1dds/stratified_estimator/