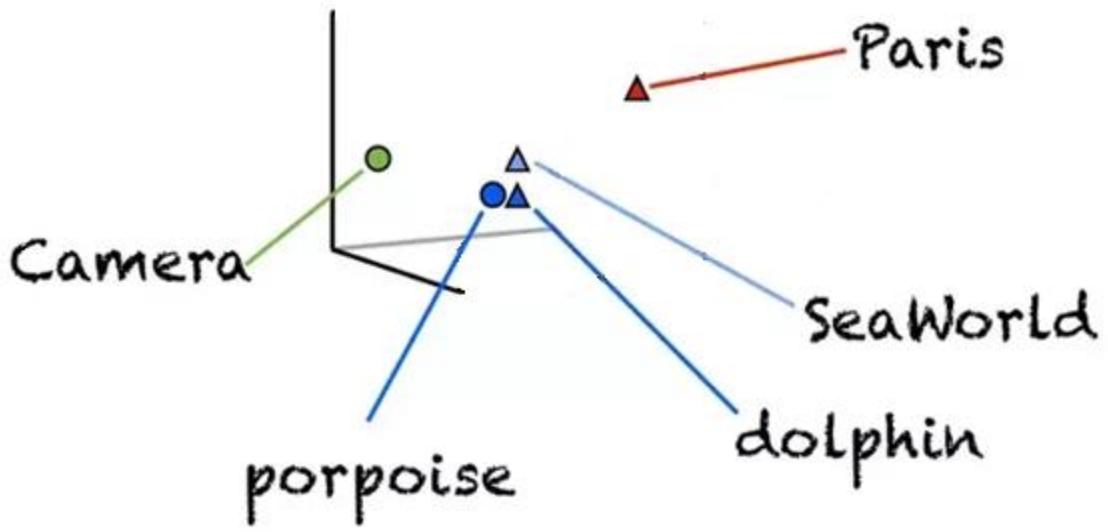
On the Downstream Performance of Compressed Word Embeddings

Avner May, Jian Zhang, Tri Dao, Chris Ré

Stanford University

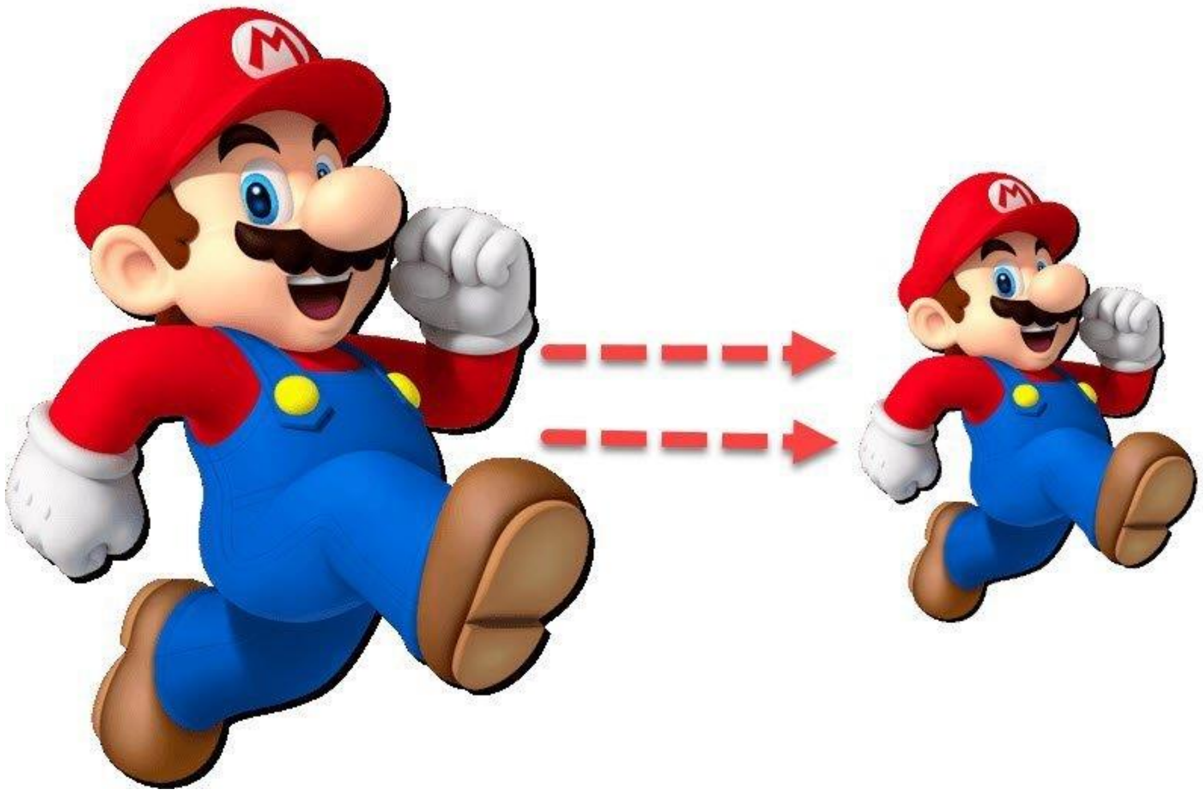
Word Embeddings



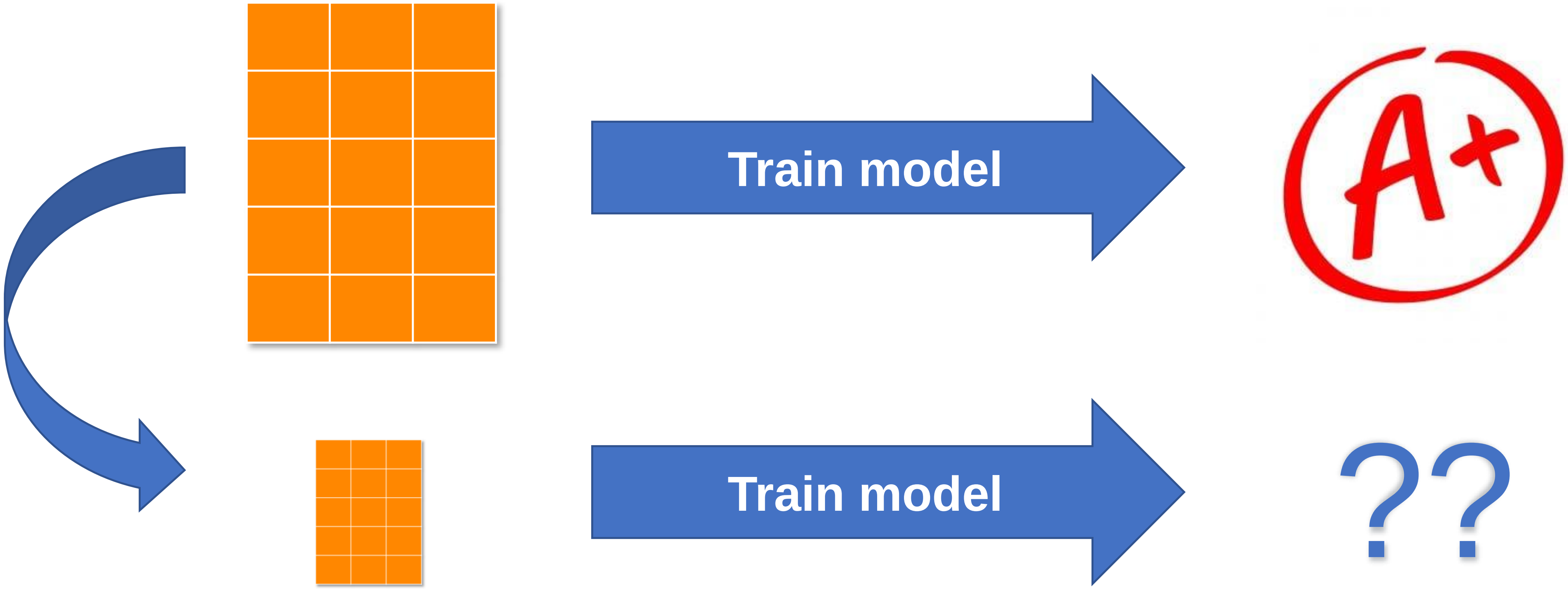
Important for strong NLP performance

Take a lot of memory

Word Embedding Compression



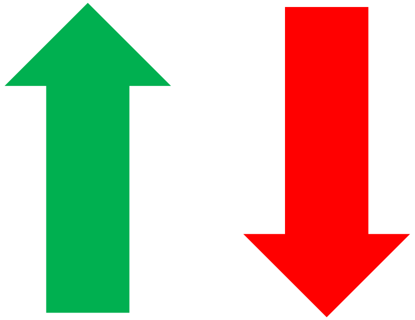
What determines whether a compressed embedding matrix will perform well on downstream tasks?



Motivating Observation

Existing ways of measuring compression quality often ***fail to explain*** relative downstream performance.

Better compression quality measure



Worse downstream performance

Our Contributions: Outline

1. Define a **new measure** of compression quality.
2. Prove **generalization bounds** using this measure.
3. Show strong **empirical correlation** w. downstream performance.
4. Use measure to **select** compressed embeddings.

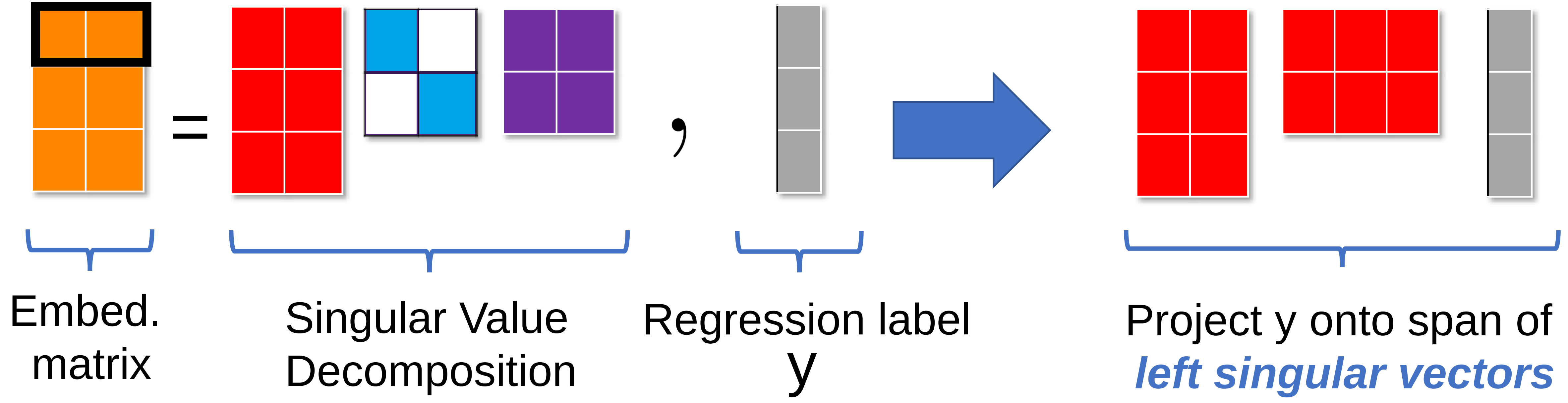
Up to *2x lower* selection error rates than the next best measure.

Defining the Measure: Intuition from Linear Regression

Observation:

Predictions are determined by the span of ***left singular vectors.***

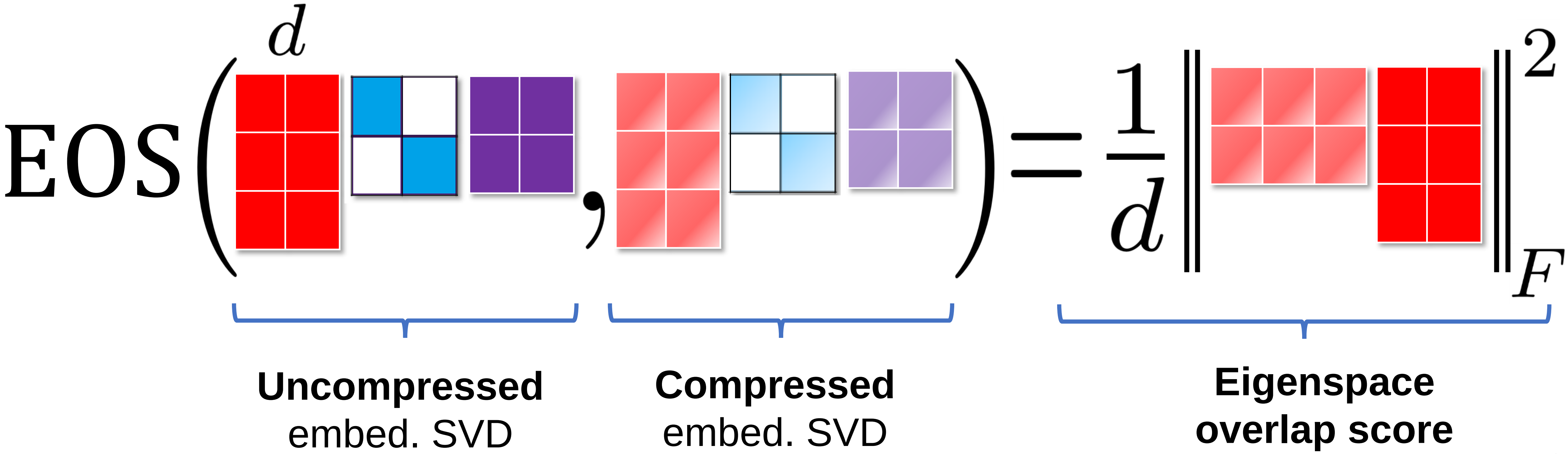
Linear regressor predictions



Defining the Measure: Eigenspace Overlap Score (EOS)

Intuition:

Measures similarity between the span of ***left singular vectors.***



Theoretical Results: Linear Regression

Theorem (informal):
Expected difference in ***test mean-squared error*** attained by ***compressed*** vs. ***uncompressed*** embeddings is ***determined by EOS***.

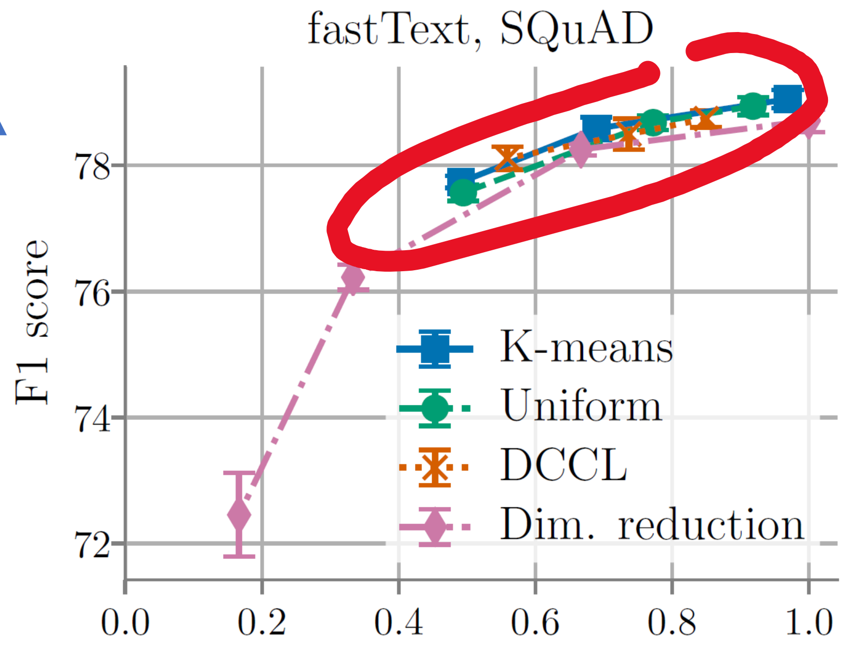
Higher EOS

Better downstream performance

Empirical Correlation: Beyond Linear Regression

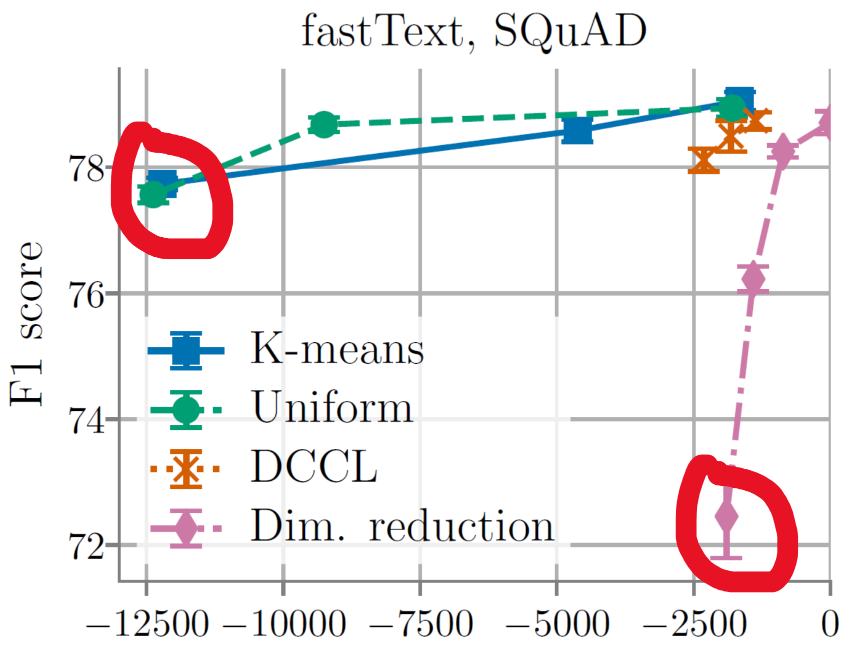
EOS attains **strong correlation** with downstream **model accuracy**.

Higher accuracy



EOS

Higher quality



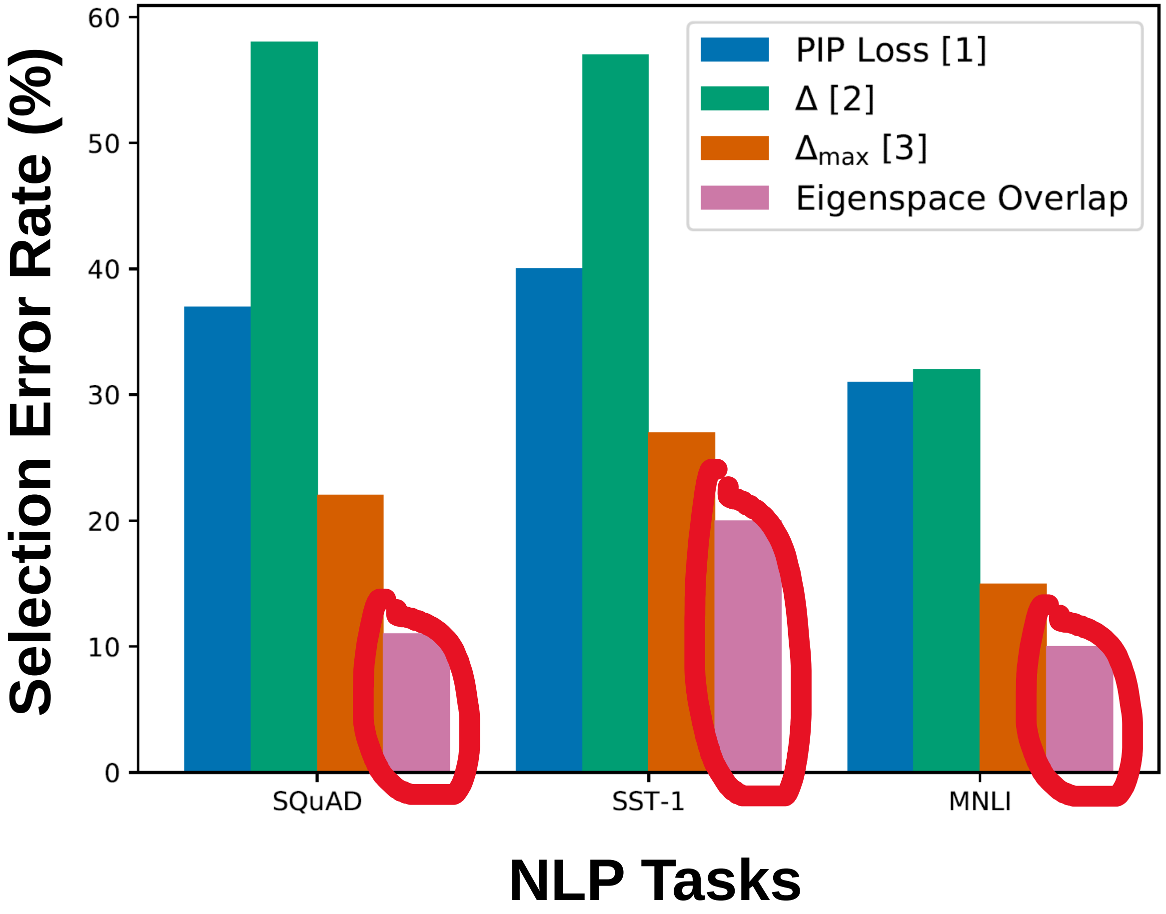
Neg. PIP Loss [1]

Higher quality

[1] Yin and Shen, *On the Dimensionality of Word Embeddings*. NeurIPS 2018.

Empirical Correlation: Beyond Linear Regression

EOS attains ***up to 2x lower selection*** error rates than 2nd best.



[1] Avron et al., ICML 2017. [2] Yin and Shen. NeurIPS 2018. [3] Zhang et al., AISTATS 2019.

Our Contributions: Summary

1. Defined a **new measure** of compression quality.
2. Proved **generalization bounds** using this measure.
3. Showed strong **empirical correlation** w. downstream perf.
4. Used measure to **select** compressed embeddings.

THANK YOU!

Poster #185, 5-7 pm Dec. 12!

Paper: https://arxiv.org/pdf/1909.01264.pdf
Code: https://github.com/HazyResearch/smallfry
E-mail: avnermay@cs.stanford.edu