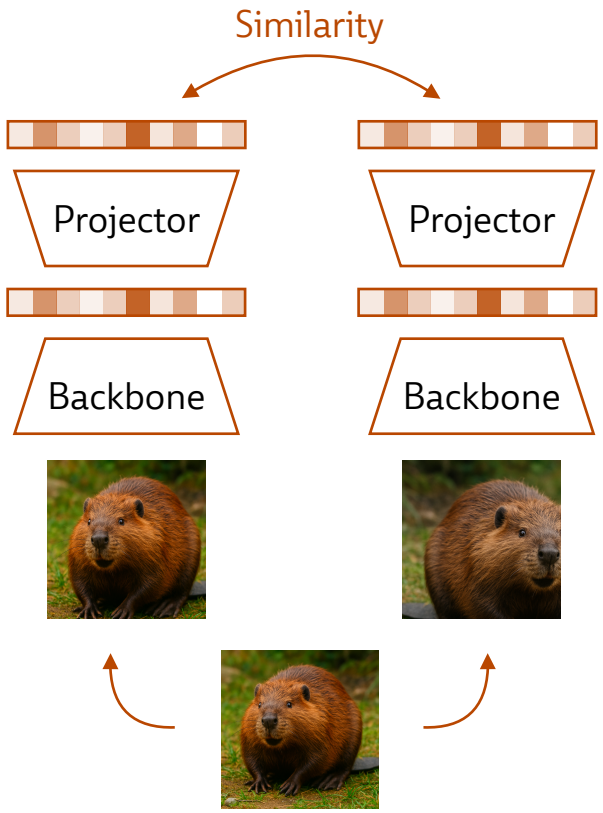
Collapse-Proof Non-Contrastive Self-Supervised Learning

Emanuele

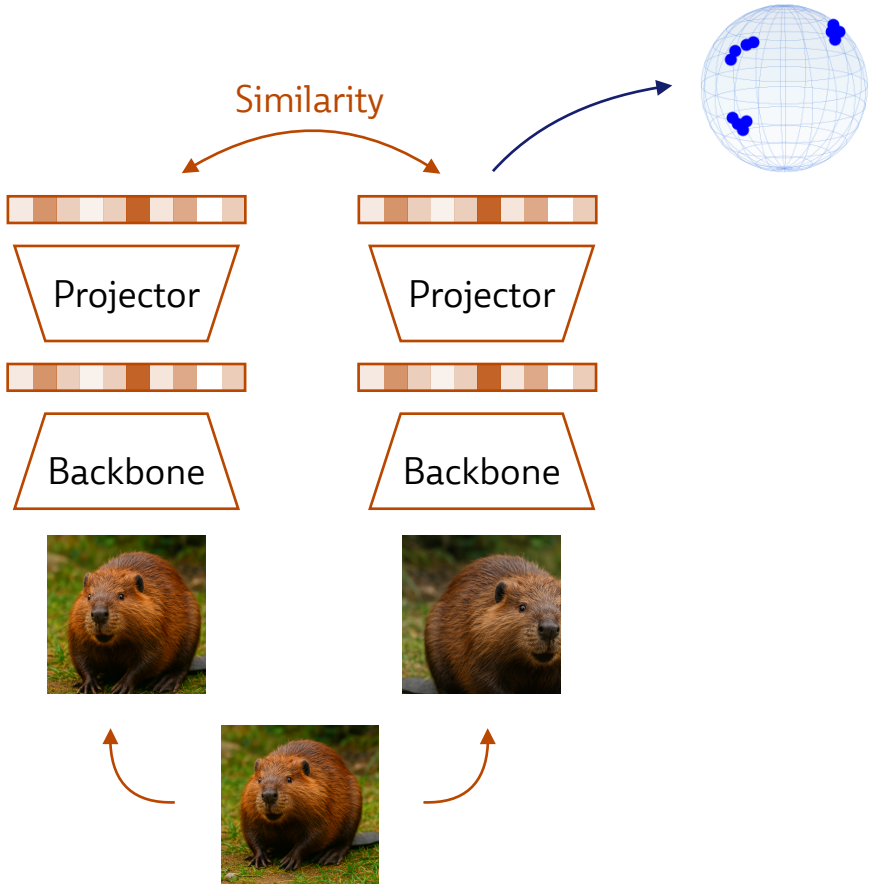
Tim

Tinne

Non-Contrastive Learning ?

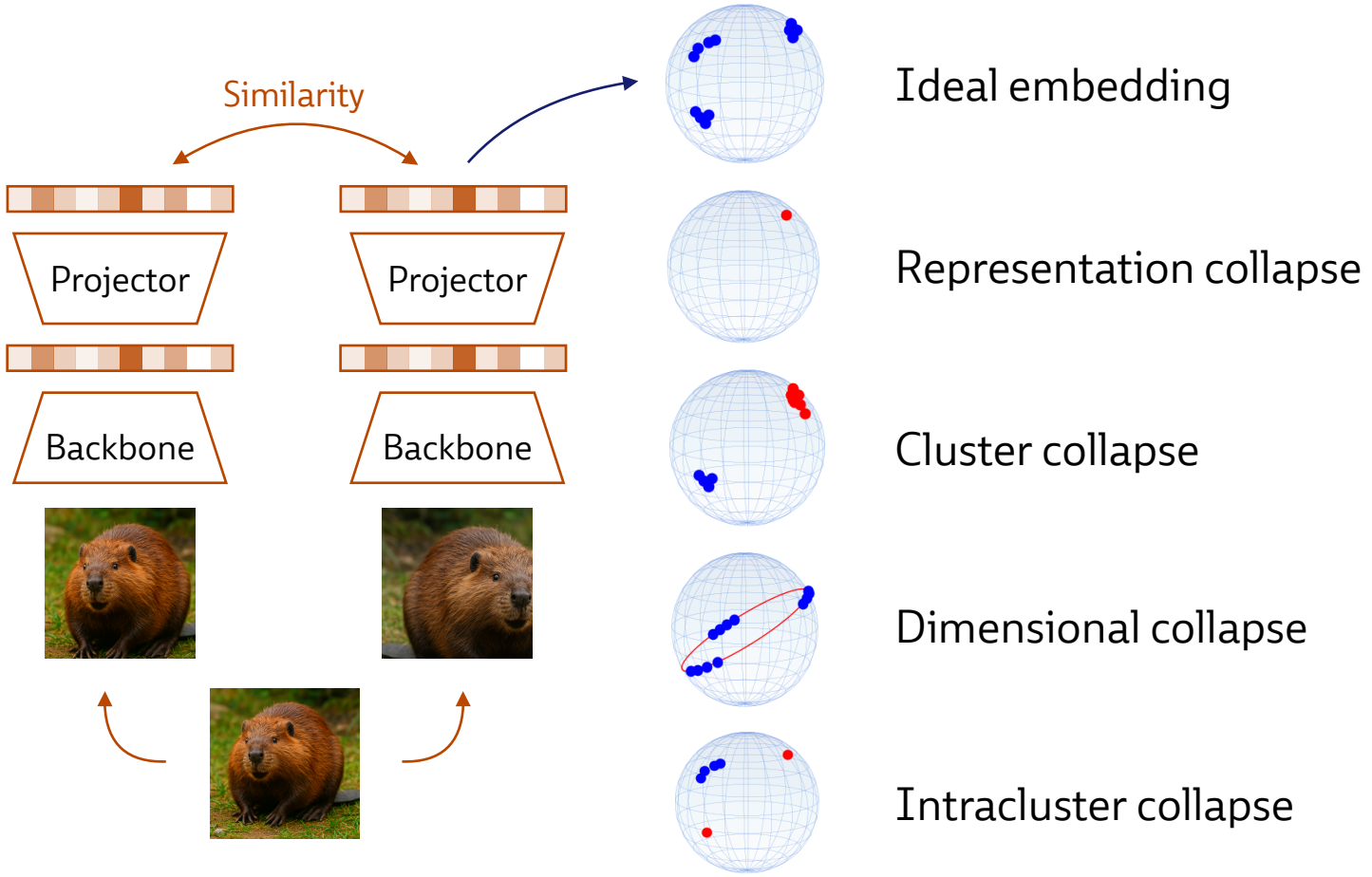


What Can Go Wrong ?



Ideal embedding

What Can Go Wrong ?



What Are The Main Non-Contrastive Strategies ?

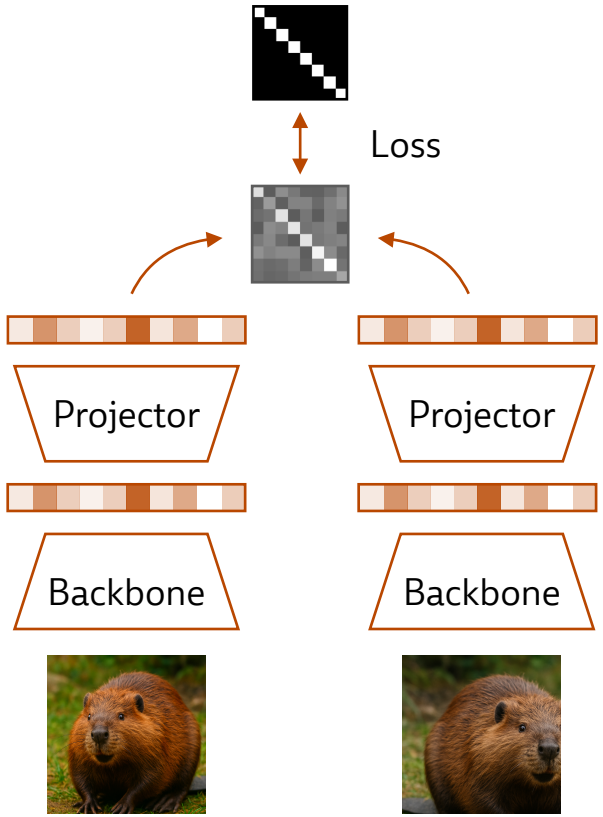
Feature Decorrelation

Barlow Twins [Zbontar et al. ICML 2021] , …

Clustering

SwAV [Caron et al. NeurIPS 2020] , …

What Are The Main Non-Contrastive Strategies ?



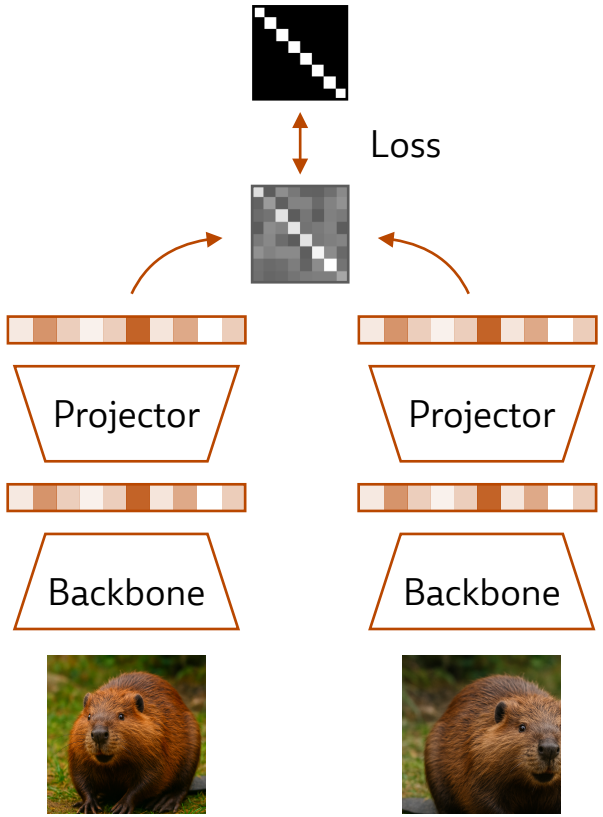
Feature Decorrelation

Barlow Twins [Zbontar et al. ICML 2021] , ...

Clustering

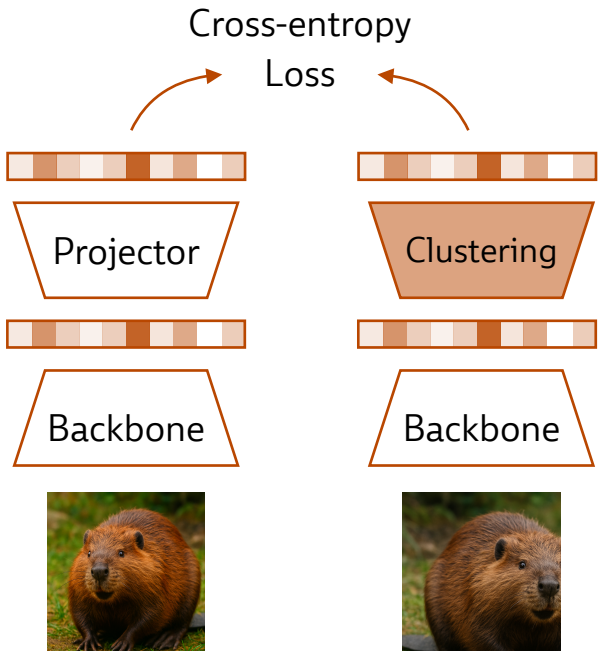
SwAV [Caron et al. NeurIPS 2020] , ...

What Are The Main Non-Contrastive Strategies ?



Feature Decorrelation

Barlow Twins [Zbontar et al. ICML 2021] , ...

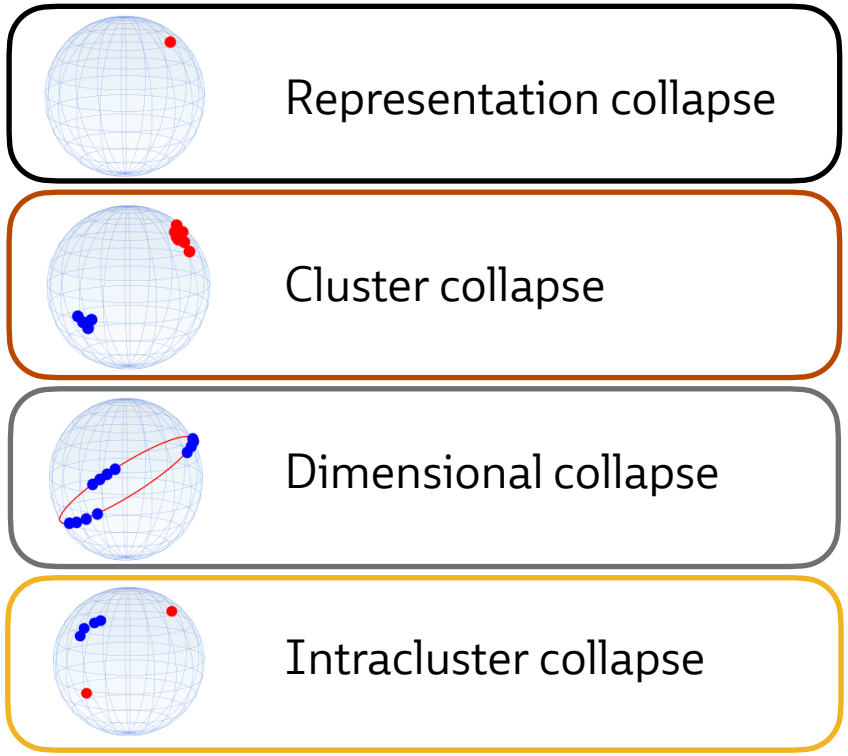


Clustering

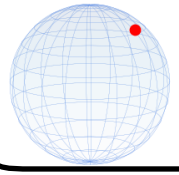
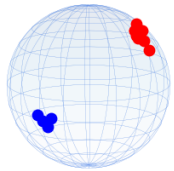
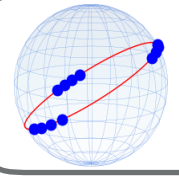
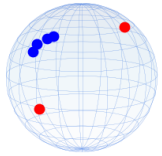
SwAV [Caron et al. NeurIPS 2020] , ...

Are Existing Strategies Collapse-Proof ?

Feature Decorrelation Clustering



Are Existing Strategies Collapse-Proof ?

	Feature Decorrelation	Clustering
 Representation collapse	✓	✓
 Cluster collapse	✗	✓
 Dimensional collapse	✓	✗
 Intracluster collapse	✗	✗

Key Contributions

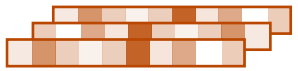
1. Sufficient conditions for avoiding collapses

Key Contributions

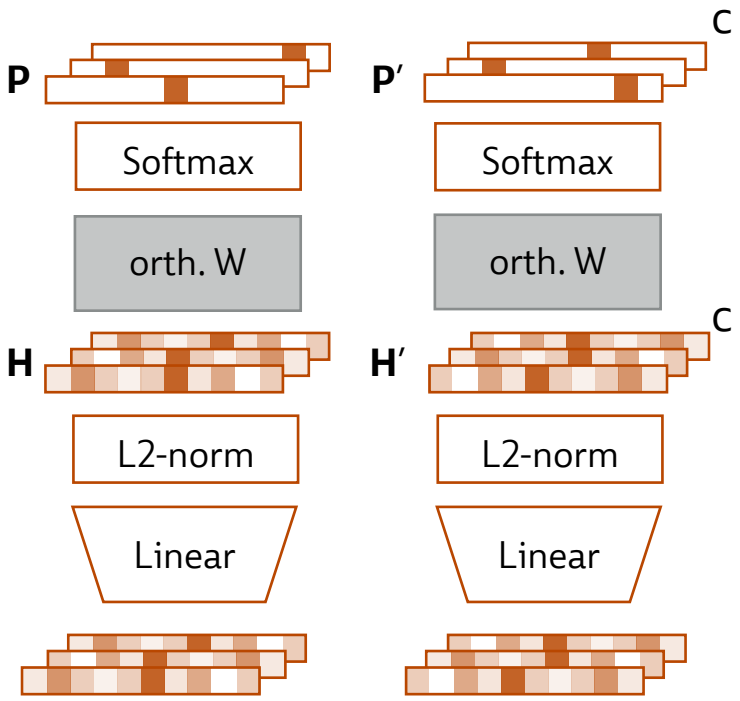
1. Sufficient conditions for avoiding collapses

2. Design guidelines for the projector and the loss function based on hyperdimensional computing

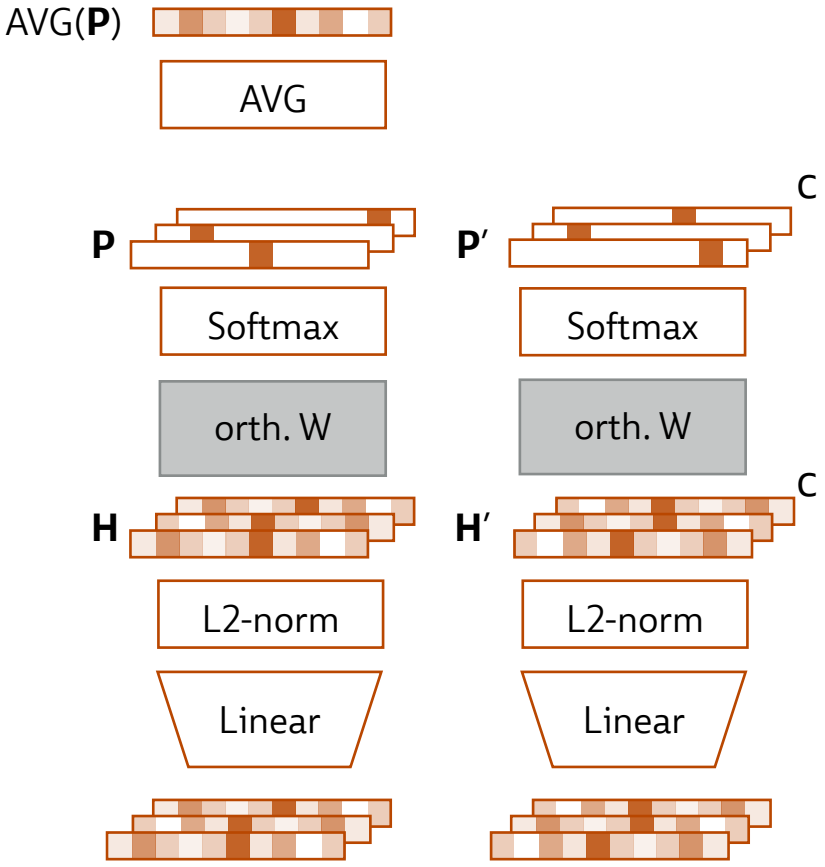
Proposed Design of Projector and Loss Function



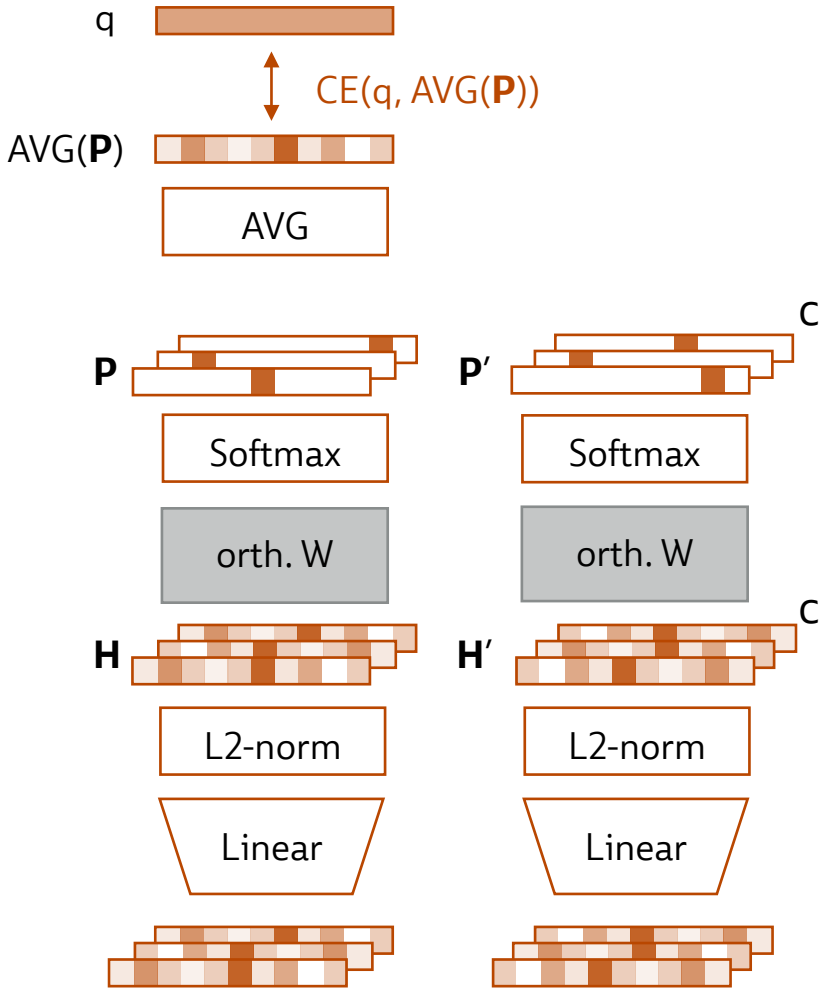
Proposed Design of Projector and Loss Function



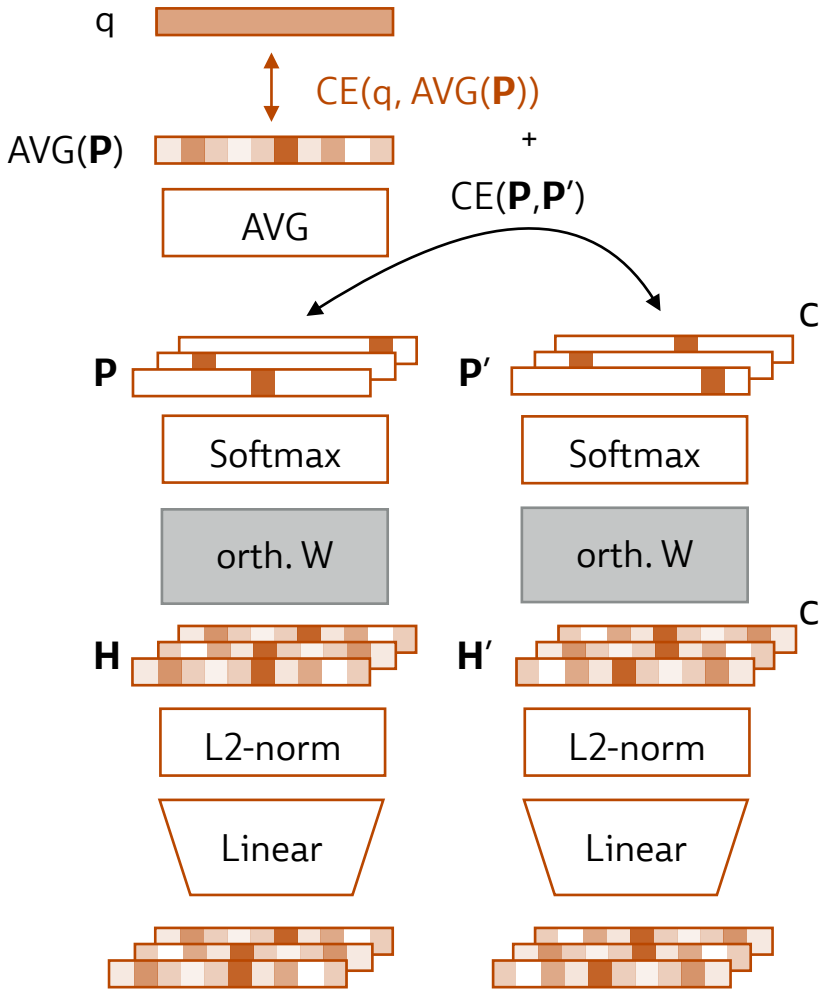
Proposed Design of Projector and Loss Function



Proposed Design of Projector and Loss Function

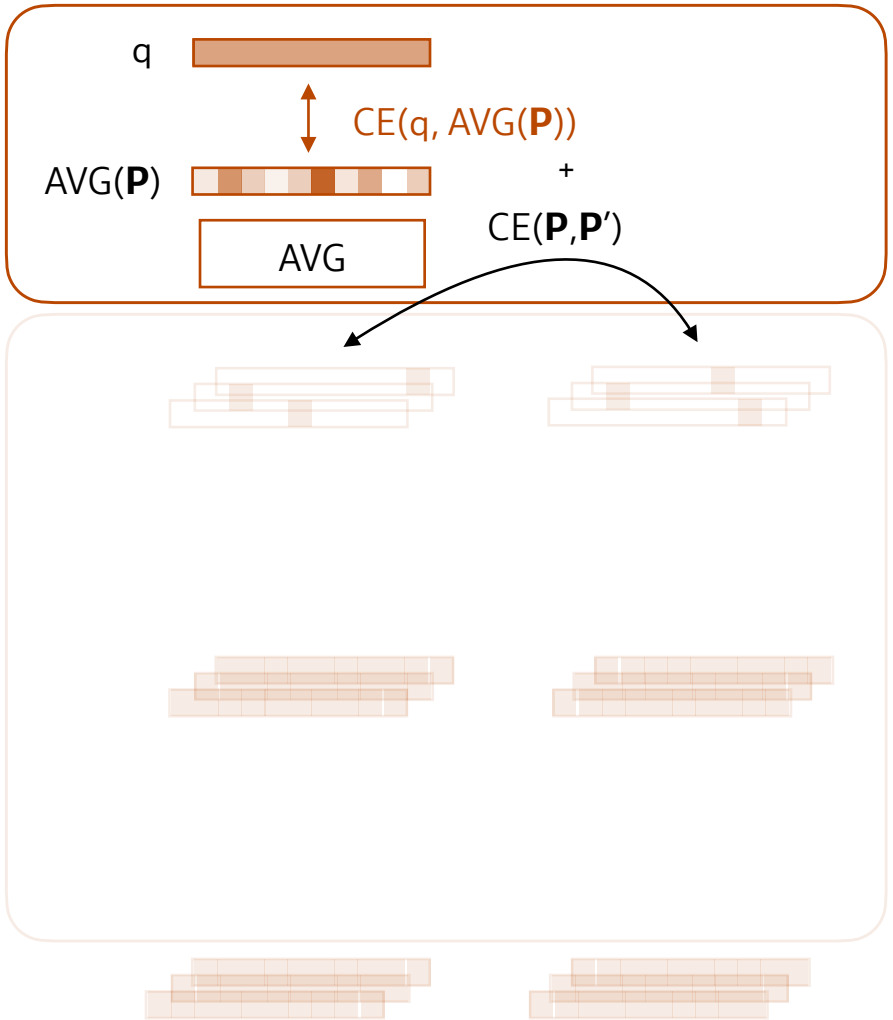


Proposed Design of Projector and Loss Function

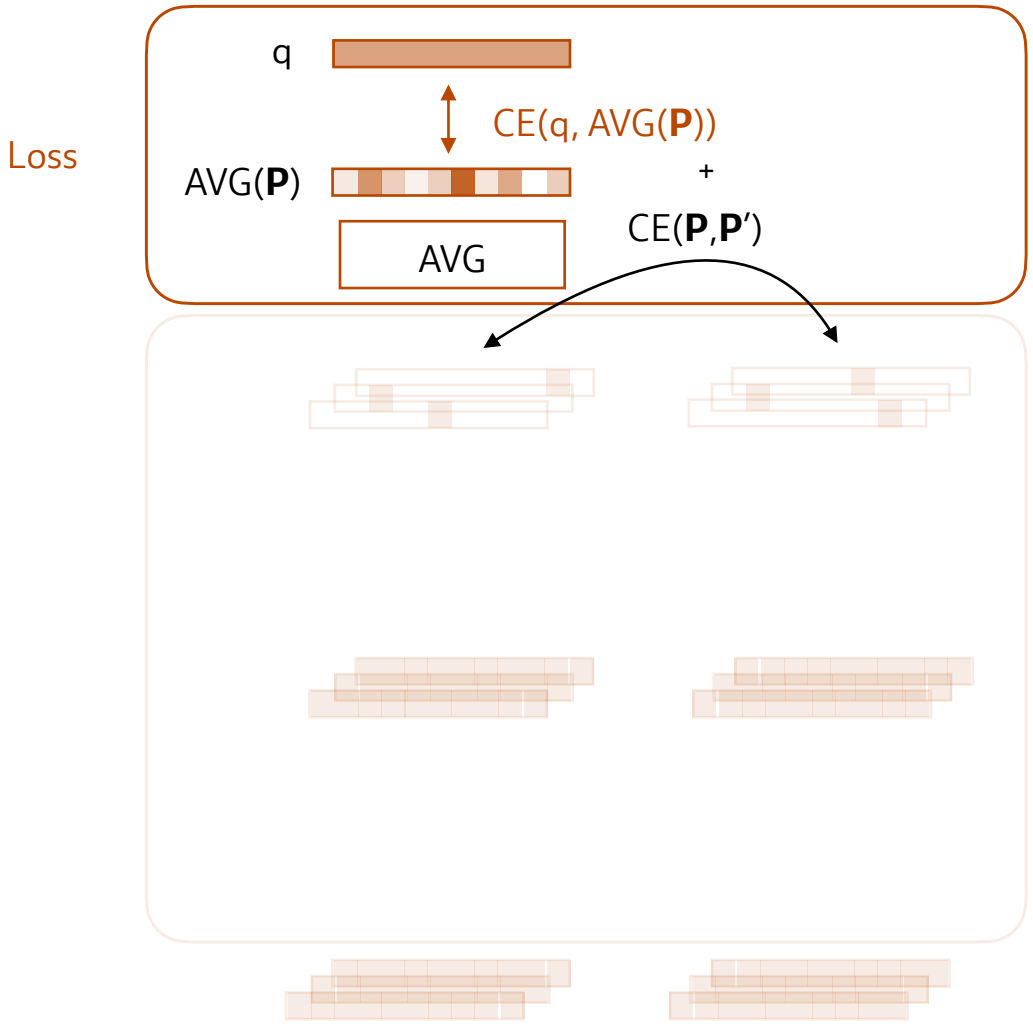


Properties of The Loss Function

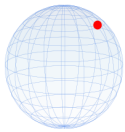
Loss



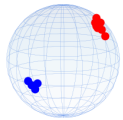
Properties of The Loss Function



(Lemma) If optima are attained, then:



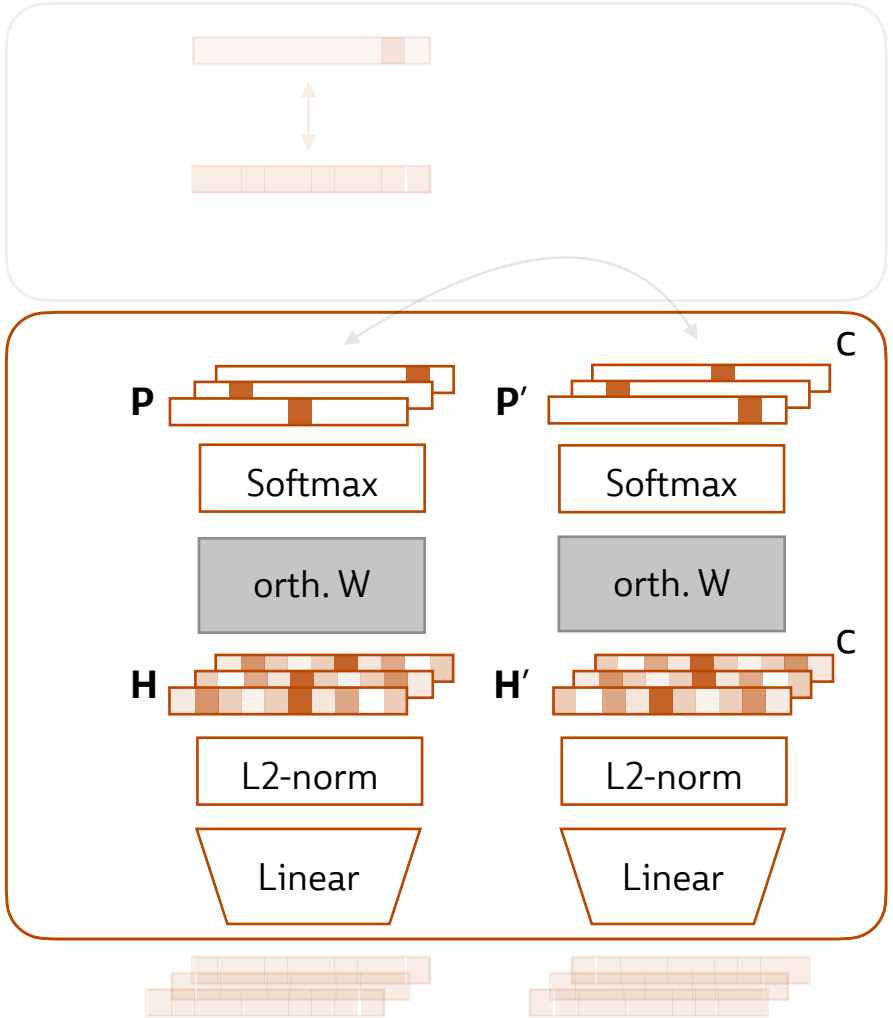
Representation collapse ✓



Cluster collapse ✓

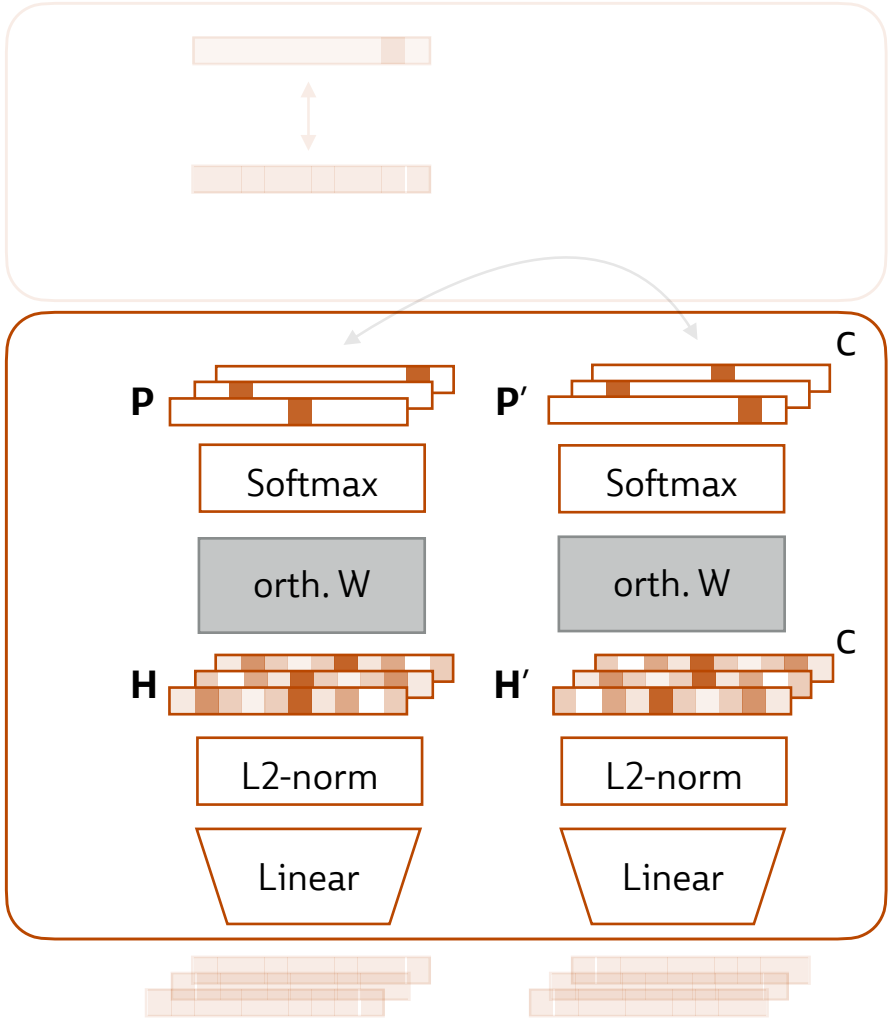
Properties of The Projector

Projector



Properties of The Projector

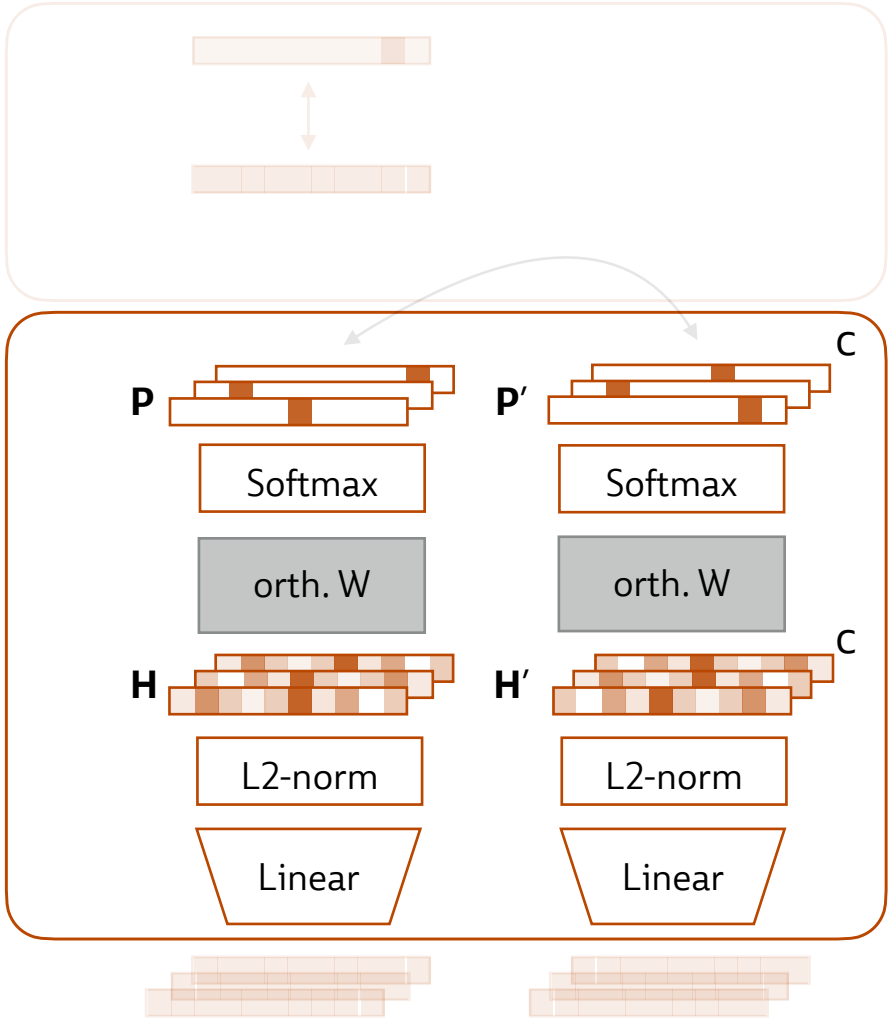
Projector



(Embedding Theorem) If optima of the loss are attained and $c \to \infty$:

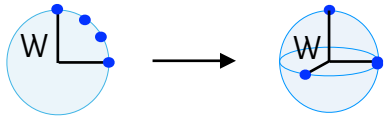
Properties of The Projector

Projector



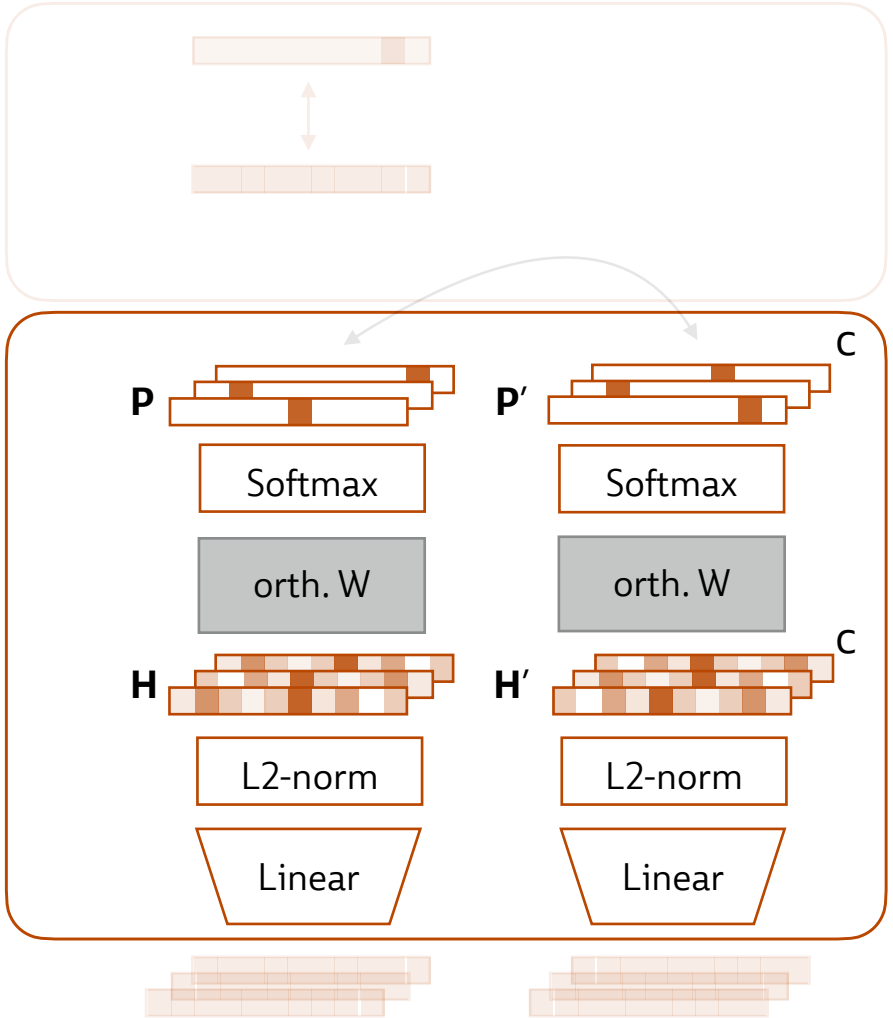
(Embedding Theorem) If optima of the loss are attained and $c \to \infty$:

1. Perfect alignment



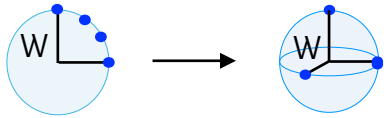
Properties of The Projector

Projector



(Embedding Theorem) If optima of the loss are attained and $c \to \infty$:

1. Perfect alignment

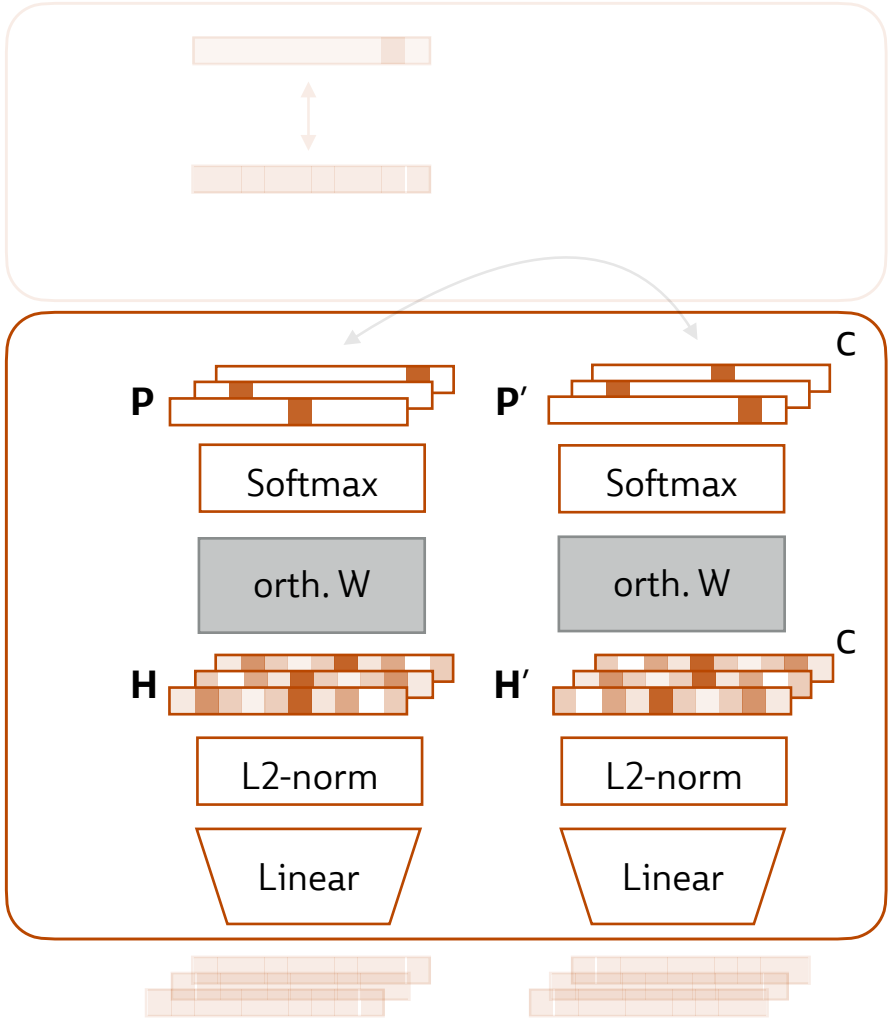


2. Sample covariance $(\mathbf{H})^T\mathbf{H}=\mathbf{I}$ is diagonal

Dimensional collapse ✓

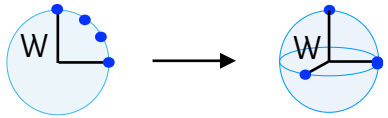
Properties of The Projector

Projector



(Embedding Theorem) If optima of the loss are attained and $c \to \infty$:

1. Perfect alignment



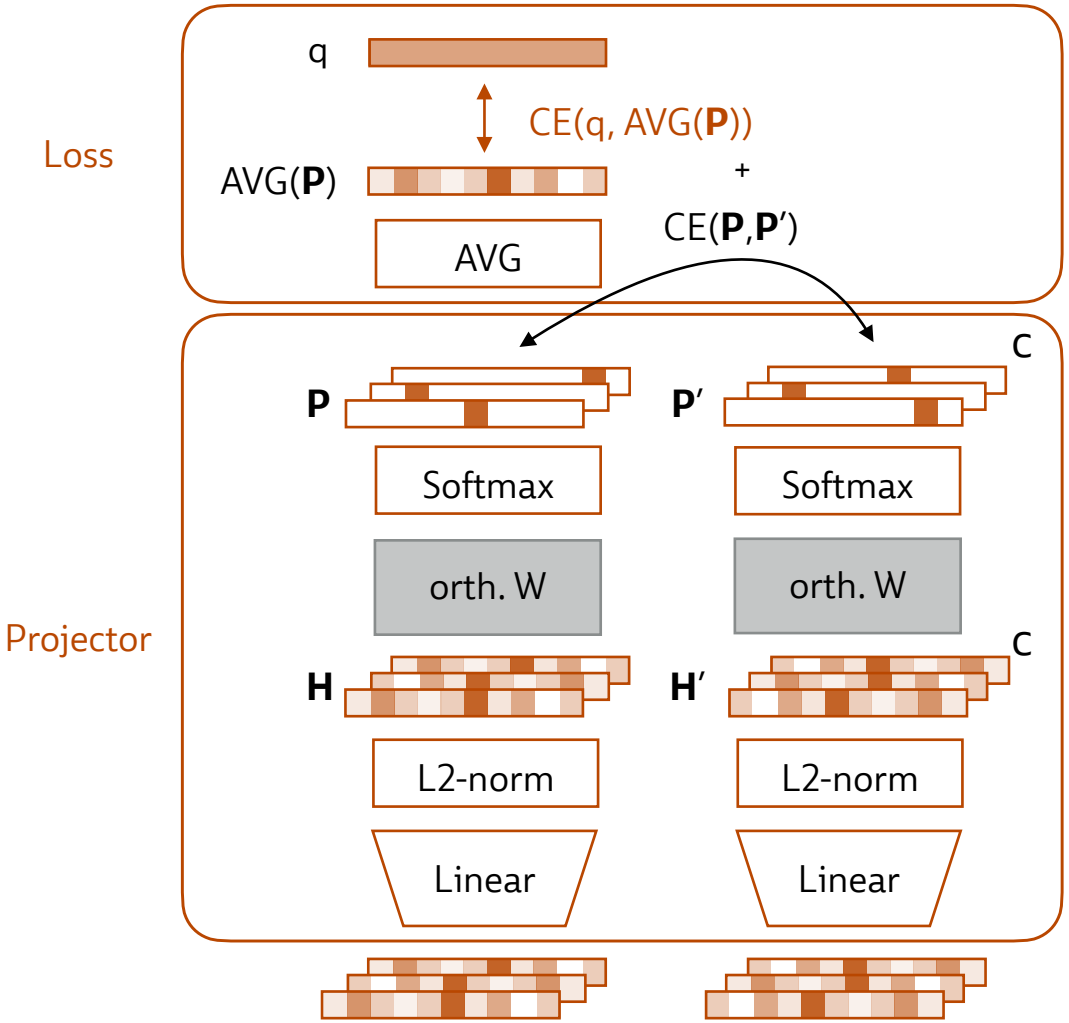
2. Sample covariance $(\mathbf{H})^T\mathbf{H}=\mathbf{I}$ is diagonal

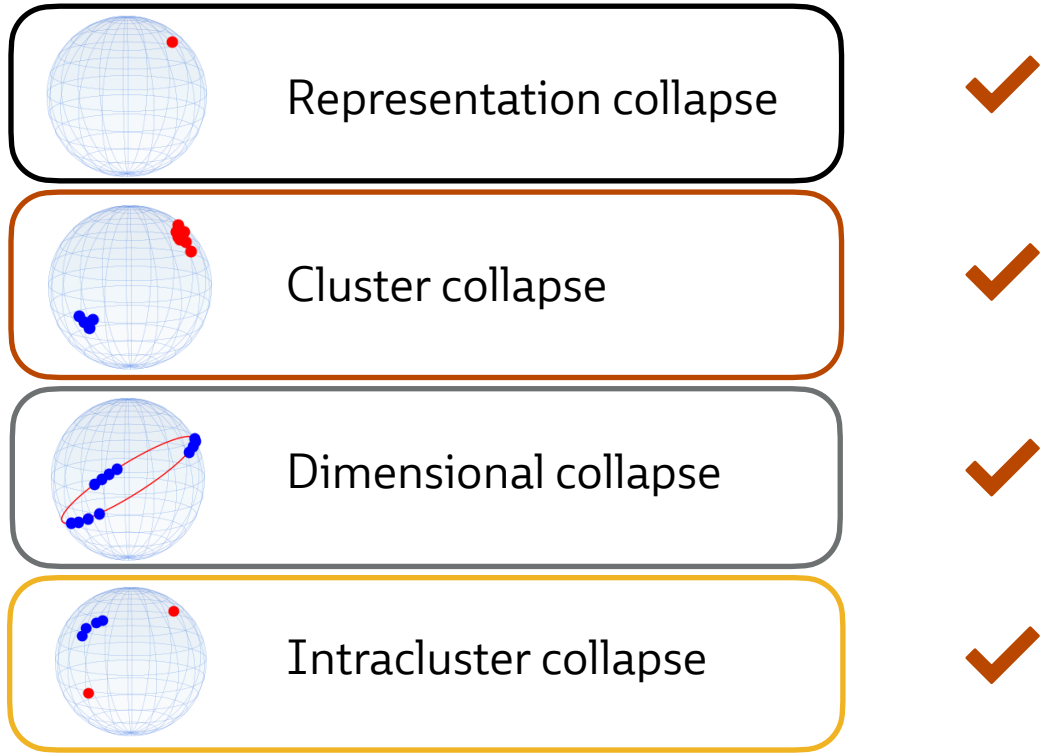
Dimensional collapse ✓

3. Adjacency $\mathbf{H}(\mathbf{H})^T$ is block-diagonal and block-size $\to 0$

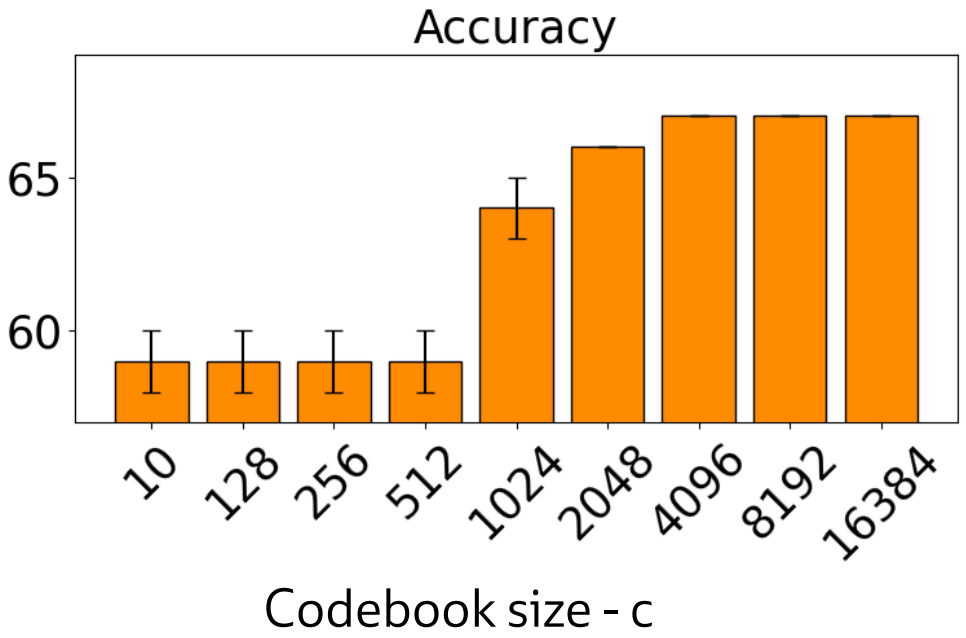
Intracluster collapse ✓

Collapse-Proof

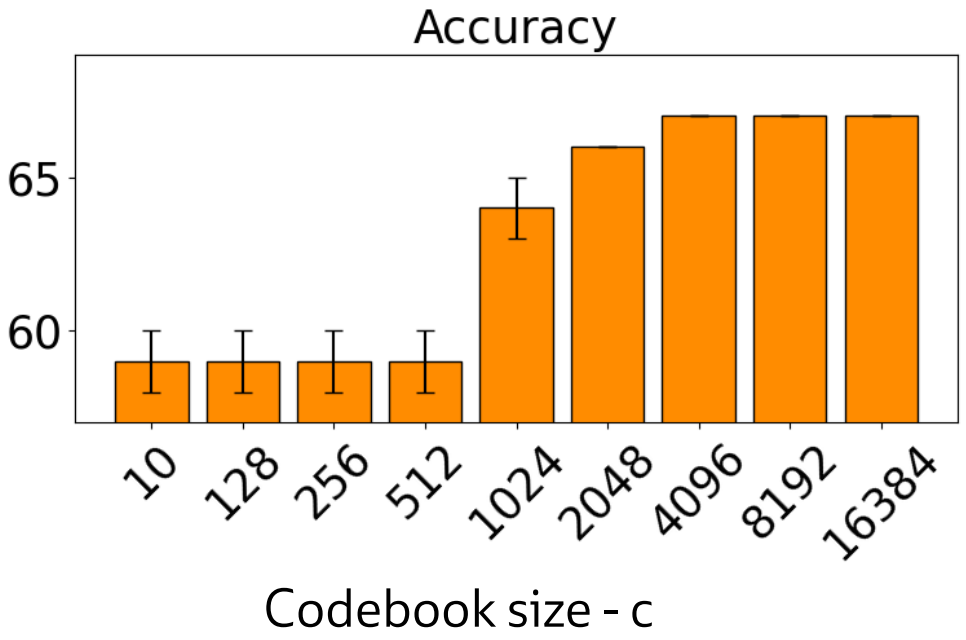


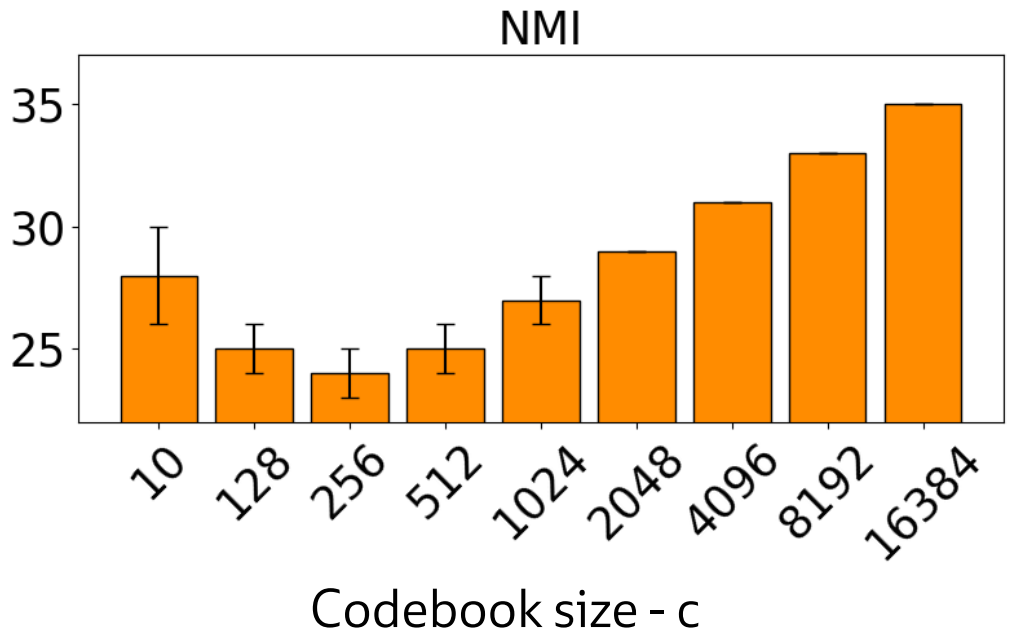


Experiments on CIFAR-10 with ResNet

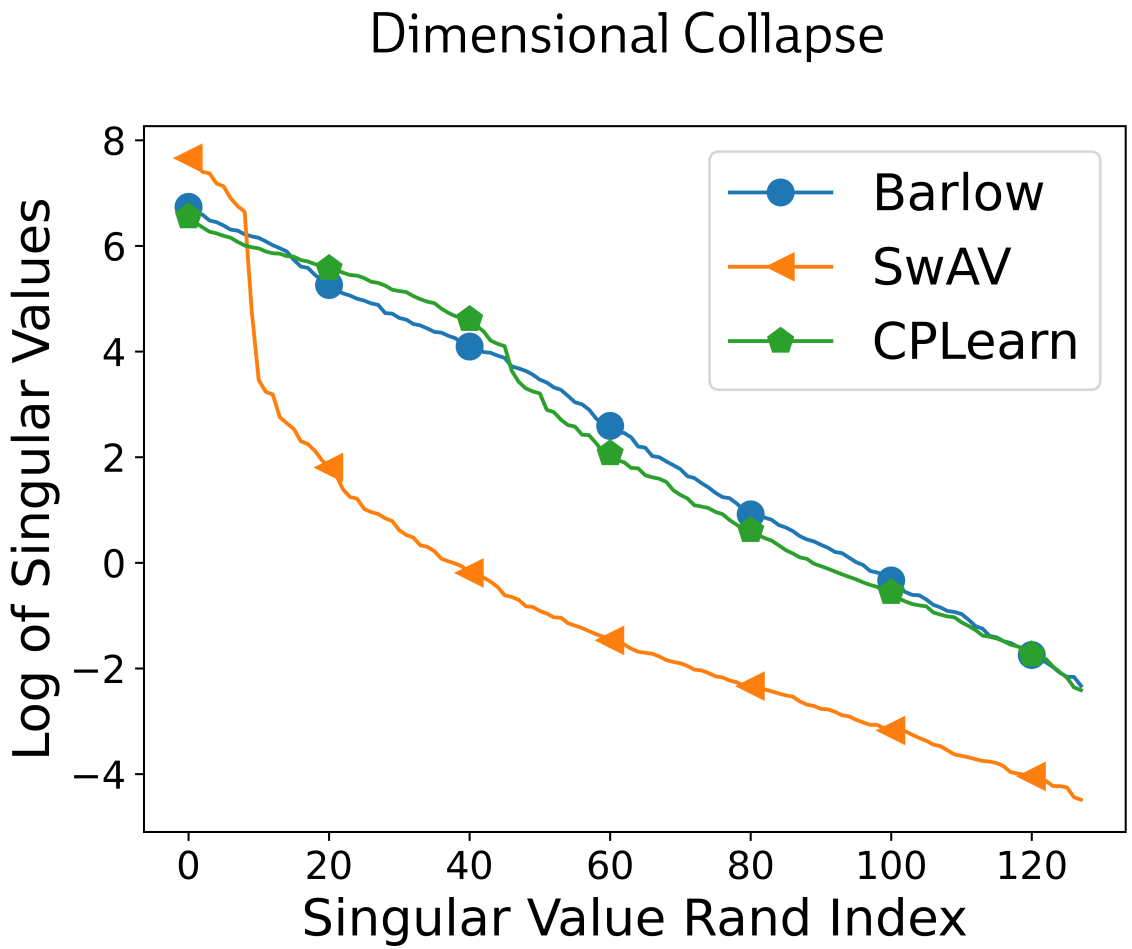


Experiments on CIFAR-10 with ResNet

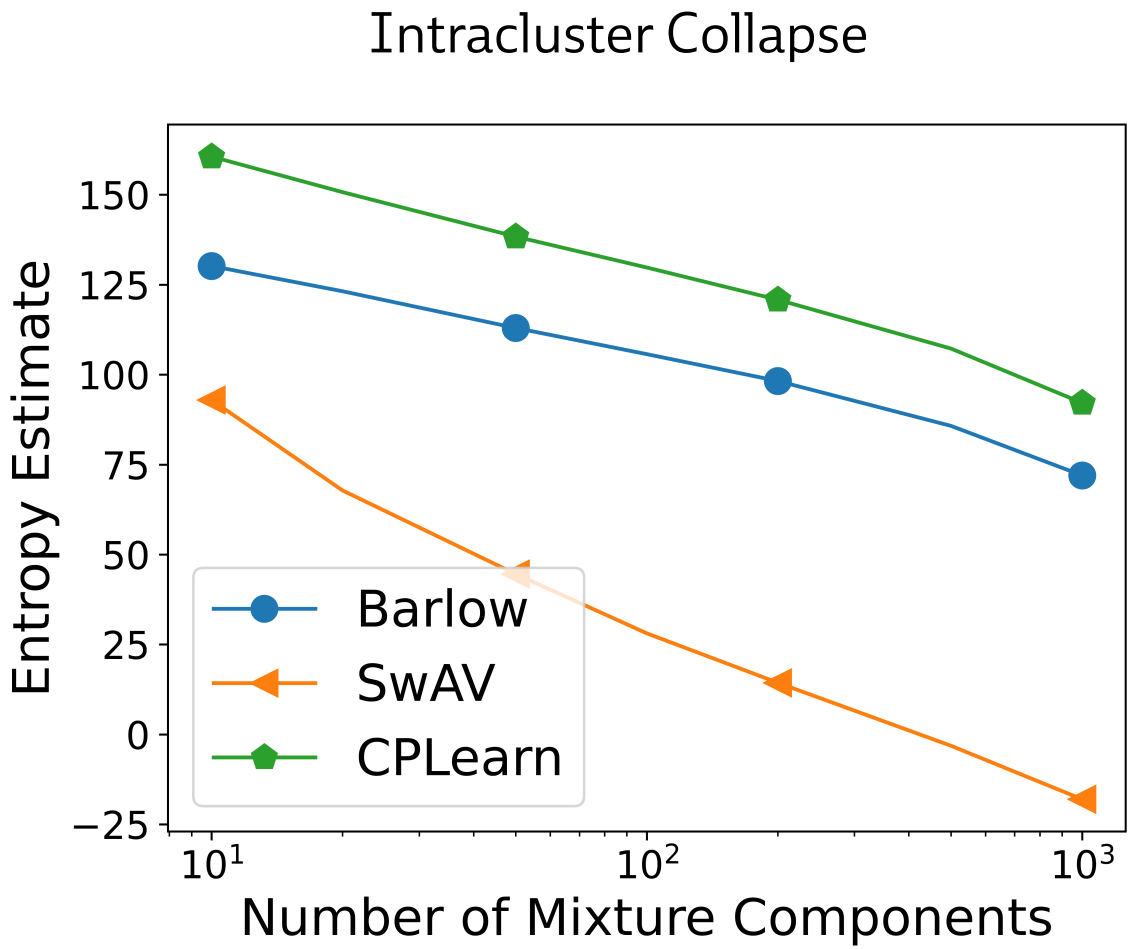




Experiments on CIFAR-10 with ResNet

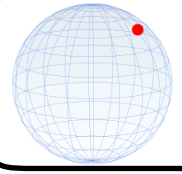
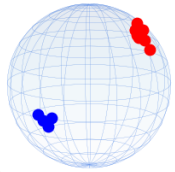
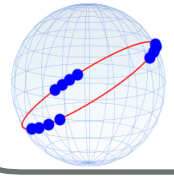
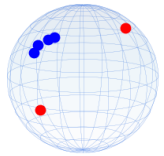


Experiments on CIFAR-10 with ResNet



Thank You !

@skiera87

	Feature Decorrelation	Clustering	CPLearn
 Representation collapse	✓	✓	✓
 Cluster collapse	✗	✓	✓
 Dimensional collapse	✓	✗	✓
 Intracluster collapse	✗	✗	✓