

Lecture 7 – Contrastive Learning

Provide short answers (max. 25 words each) to the following questions.

- What is the main goal of Contrastive Learning?
- What are “anchor,” “positive,” and “negative” samples in contrastive learning?
- How does contrastive learning differ from supervised learning?
- Why is contrastive learning useful in domains with scarce labels?
- What property of representations does contrastive learning aim to achieve?
- What is the key idea behind contrastive loss?
- What role does the margin parameter play in contrastive loss?
- What is the embedding function in contrastive learning?
- What is triplet loss?
- What does N-pair loss improve over triplet loss?
- What is Noise-Contrastive Estimation (NCE)?
- What additional components does Noise-Contrastive Estimation (NCE) introduce?
- What is InfoNCE?
- Why is cosine similarity often used in Contrastive Learning?
- What advantage does cosine similarity have over Euclidean distance?
- What is SimCLR?
- What is the role of the projection head in SimCLR?
- What is Momentum Contrast (MoCo)?
- What is the key innovation in MoCo?
- What is Bootstrap Your Own Latent (BYOL)?
- What is SimSiam?
- Why are data augmentations crucial in contrastive learning?
- Why is a nonlinear projection head important?
- How does model size affect contrastive learning performance?
- What is the effect of temperature scaling in contrastive loss?
- Does Contrastive Learning benefit from longer training?
- Can small batch sizes work in Contrastive Learning?
- Why might dataset bias affect Contrastive Learning?

Multiple-choice questions

What is the main goal of Contrastive Learning?

- A) Predict labels directly
- B) Differentiate similar and dissimilar pairs
- C) Reduce model size
- D) Increase batch size

Which pairs are pulled closer in Contrastive Learning?

- A) Negative pairs
- B) Positive pairs
- C) Random pairs
- D) Augmented negatives

Which pairs are pushed apart in Contrastive Learning?

- A) Negative pairs
- B) Positive pairs
- C) Augmented positives
- D) Random pairs

What is an anchor in Contrastive Learning?

- A) A negative sample
- B) A reference sample
- C) A projection head
- D) A loss function

What is the main advantage of self-supervised Contrastive Learning?

- A) Requires large labeled datasets
- B) Uses only supervised loss
- C) Works without explicit labels
- D) Eliminates augmentations

What does the margin parameter in contrastive loss control?

- A) Batch size
- B) Temperature scaling
- C) Learning rate
- D) Minimum distance between negative pairs

What is triplet loss designed to do?

- A) Push anchor and positive apart
- B) Pull anchor and positive together, push anchor and negative apart
- C) Remove negative samples
- D) Normalize embeddings

What does N-pair loss improve over triplet loss?

- A) Uses fewer negatives
- B) Removes positives
- C) Uses multiple negatives per anchor
- D) Eliminates embeddings

What is Noise-Contrastive Estimation (NCE)?

- A) Treats each instance as its own class
- B) Uses only positive pairs
- C) Removes normalization
- D) Requires labeled data

Which similarity metric is commonly used in Contrastive Learning?

- A) Euclidean distance
- B) Cosine similarity
- C) Manhattan distance
- D) KL divergence

Why is Cosine similarity preferred?

- A) Captures magnitude only
- B) Works only for labeled data
- C) Requires normalization
- D) Captures orientation regardless of magnitude

What is SimCLR?

- A) A supervised learning framework
- B) A contrastive learning framework using augmentations
- C) A convolutional network
- D) A recurrent network

What is the role of the projection head in SimCLR?

- A) Removes embeddings
- B) Computes attention weights
- C) Maps features to a latent space for contrastive loss
- D) Normalizes inputs

What is the key innovation in MoCo?

- A) Removes negative samples
- B) Maintains a dynamic queue of negatives
- C) Uses only positive pairs
- D) Eliminates augmentations

What is SimSiam's main feature?

- A) Requires large batches
- B) Uses momentum encoders
- C) Learns without negatives using stop-gradient

D) Requires labeled data

Why are data augmentations crucial in Contrastive Learning?

- A) They reduce batch size
- B) They create diverse positive pairs
- C) They eliminate negatives
- D) They normalize embeddings

Why is a nonlinear projection head important?

- A) Improves representation quality before final layer
- B) Reduces model size
- C) Removes augmentations
- D) Normalizes embeddings

How does model size affect Contrastive Learning?

- A) Larger models perform worse
- B) Larger models improve accuracy and reduce label needs
- C) Model size has no effect
- D) Smaller models always outperform

What does temperature scaling in contrastive loss control?

- A) Learning rate
- B) Margin parameter
- C) Batch size
- D) Sharpness of similarity distribution

Does contrastive learning benefit from longer training?

- A) No
- B) Yes, improves representation quality
- C) Only for supervised tasks
- D) Only with large batches

Can small batch sizes work in contrastive learning?

- A) No
- B) Yes, with proper hyperparameter tuning
- C) Only with negatives removed
- D) Only for supervised learning

Why might dataset bias affect Contrastive Learning?

- A) Makes training harder
- B) Eliminates negative pairs

- C) Simplifies positive pair generation
- D) Improves projection head performance

What is the main idea behind InfoNCE?

- A) Predict labels directly
- B) Use only negative pairs
- C) Remove augmentations
- D) Maximize agreement between transformed representations

Which framework introduced momentum encoder?

- A) SimCLR
- B) MoCo
- C) BYOL
- D) SimSiam

Which method does NOT require negative pairs?

- A) SimCLR
- B) MoCo
- C) BYOL
- D) NCE

Which method uses stop-gradient to prevent collapse?

- A) SimCLR
- B) SimSiam
- C) MoCo
- D) InfoNCE