Contrastive representation learning

36-708

Gabriel Moryoussef

April 25th 2023

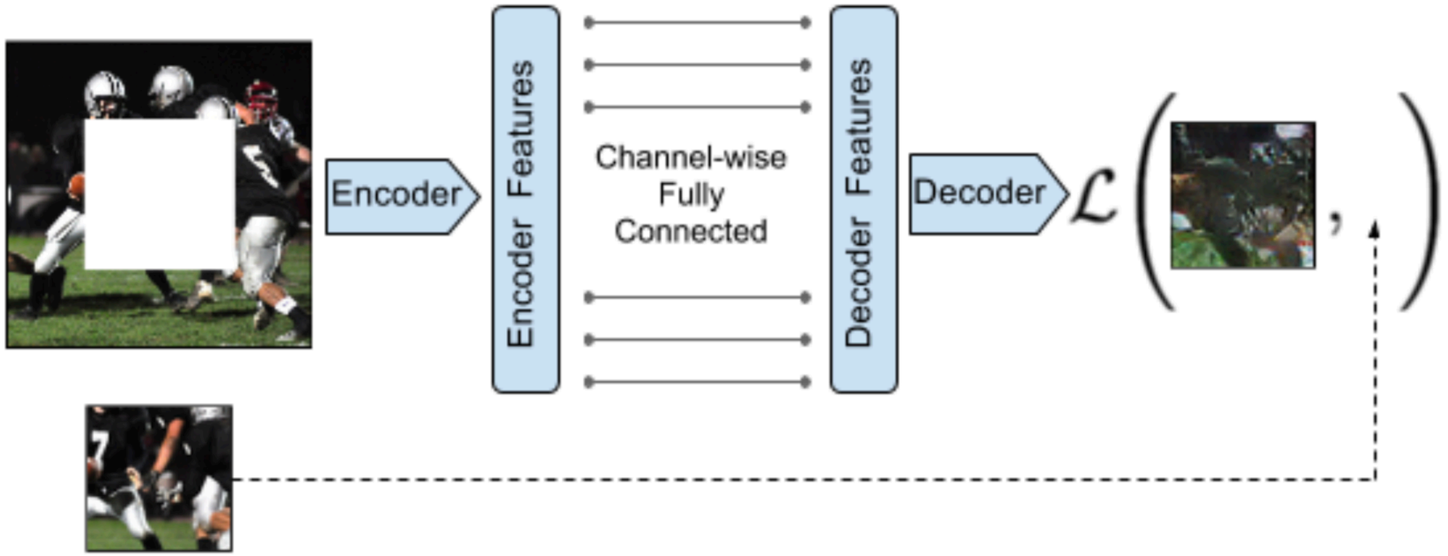
What is contrastive learning?

Goal/definition: Learn an **embedding** space in which **similar** sample pairs stay **close** to each other while dissimilar ones are far apart.

Works in supervised and self-supervised learning

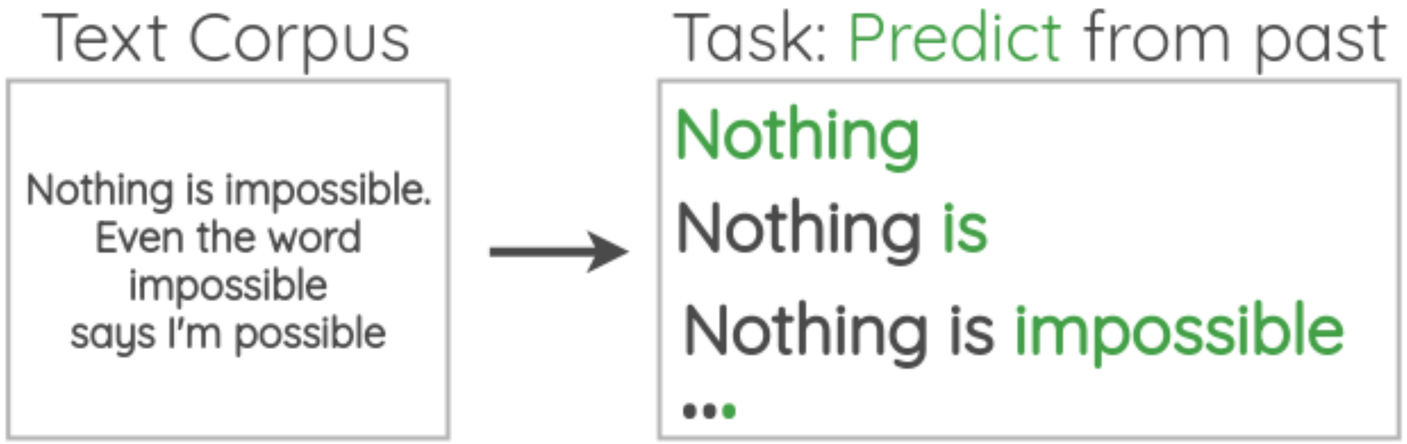
Recall on self-supervised learning

- Idea: create labels from inputs



Pathak et al. 2016

Other example: Language embedding models, word2vec…



Contrastive losses

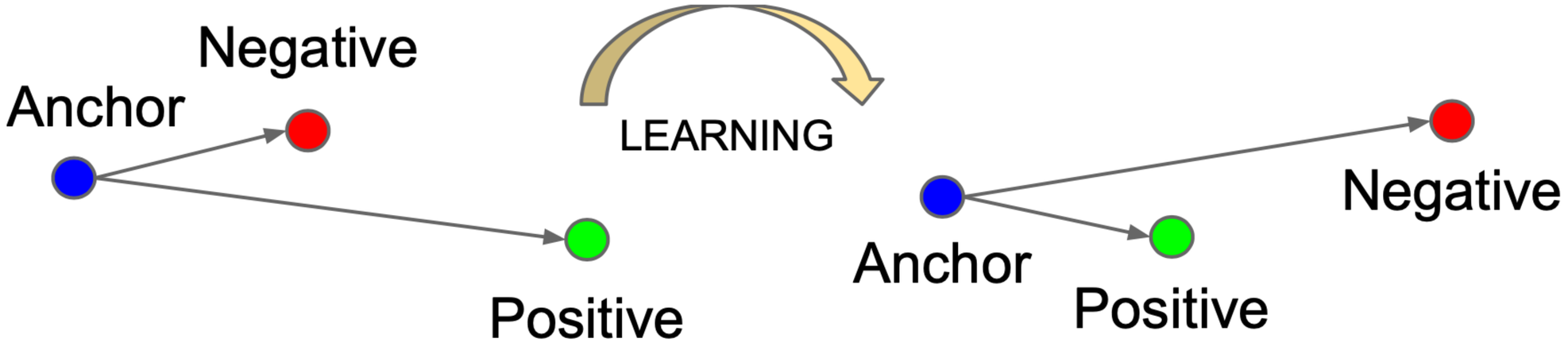
- Contrastive loss

$$\mathscr{L}_{\text{cont}}\left(\mathbf{x}_i, \mathbf{x}_j, \theta\right) = \mathbf{1}\left[y_i = y_j\right] \left\| f_\theta\left(\mathbf{x}_i\right) - f_\theta\left(\mathbf{x}_j\right) \right\|_2^2 + \mathbf{1}\left[y_i \neq y_j\right] \max\left(0, \epsilon - \left\| f_\theta\left(\mathbf{x}_i\right) - f_\theta\left(\mathbf{x}_j\right) \right\|_2\right)^2$$

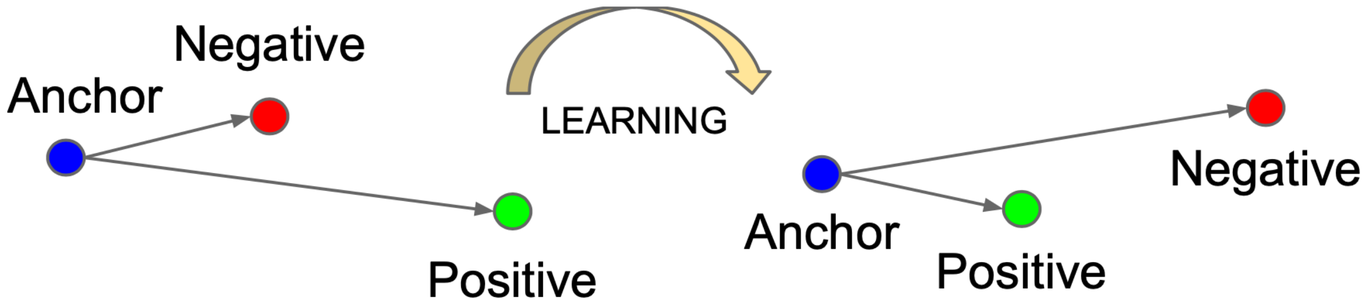
(Chopra et al. 2005)

! ! Tends to map similar inputs to the same point

- Triplet loss



(Schroff et al. 2015)



$$\mathscr{L}_{\text{triplet}}\left(\mathbf{x}, \mathbf{x}^{+}, \mathbf{x}^{-}\right) = \max\left(0, \left\| f(\mathbf{x}) - f\left(\mathbf{x}^{+}\right) \right\|_2^2 - \left\| f(\mathbf{x}) - f\left(\mathbf{x}^{-}\right) \right\|_2^2 + \epsilon\right)$$

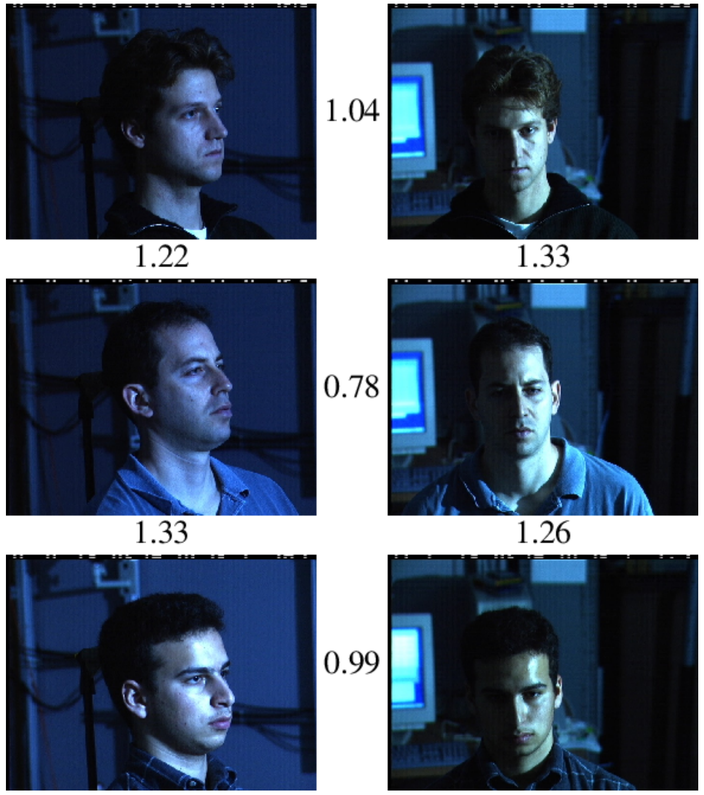
Easy generalization: the N-pair loss

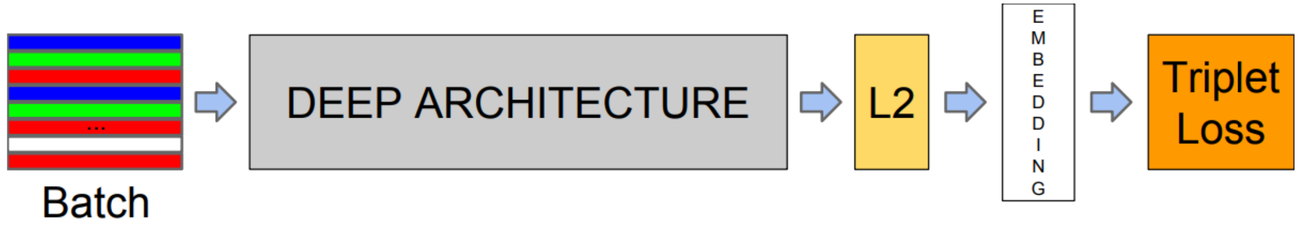
$$\mathscr{L}_{\text{N-pair}}\left(\mathbf{x}, \mathbf{x}^{+}, \left\{\mathbf{x}_i^{-}\right\}_{i=1}^{N-1}\right) = \log\left(1 + \sum_{i=1}^{N-1} \exp\left(f(\mathbf{x})^{\top} f\left(\mathbf{x}_i^{-}\right) - f(\mathbf{x})^{\top} f\left(\mathbf{x}^{+}\right)\right)\right)$$

$$= -\log \frac{\exp\left(f(\mathbf{x})^{\top} f\left(\mathbf{x}^{+}\right)\right)}{\exp\left(f(\mathbf{x})^{\top} f\left(\mathbf{x}^{+}\right)\right) + \sum_{i=1}^{N-1} \exp\left(f(\mathbf{x})^{\top} f\left(\mathbf{x}_i^{-}\right)\right)}$$

(Sohn et al. 2016)

Example: FaceNet





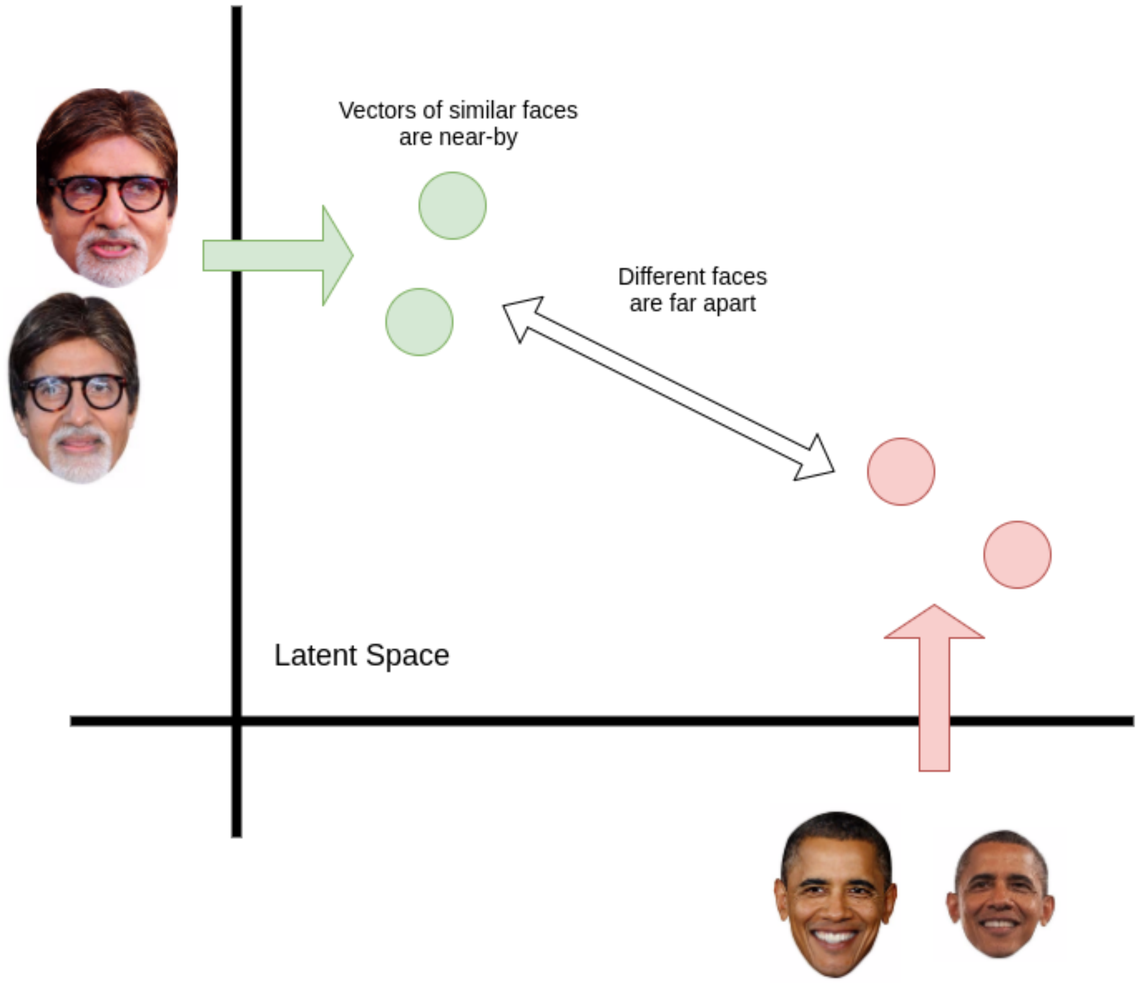
Embedding dimension = 128

(Schroff et al. 2015)

Benefits

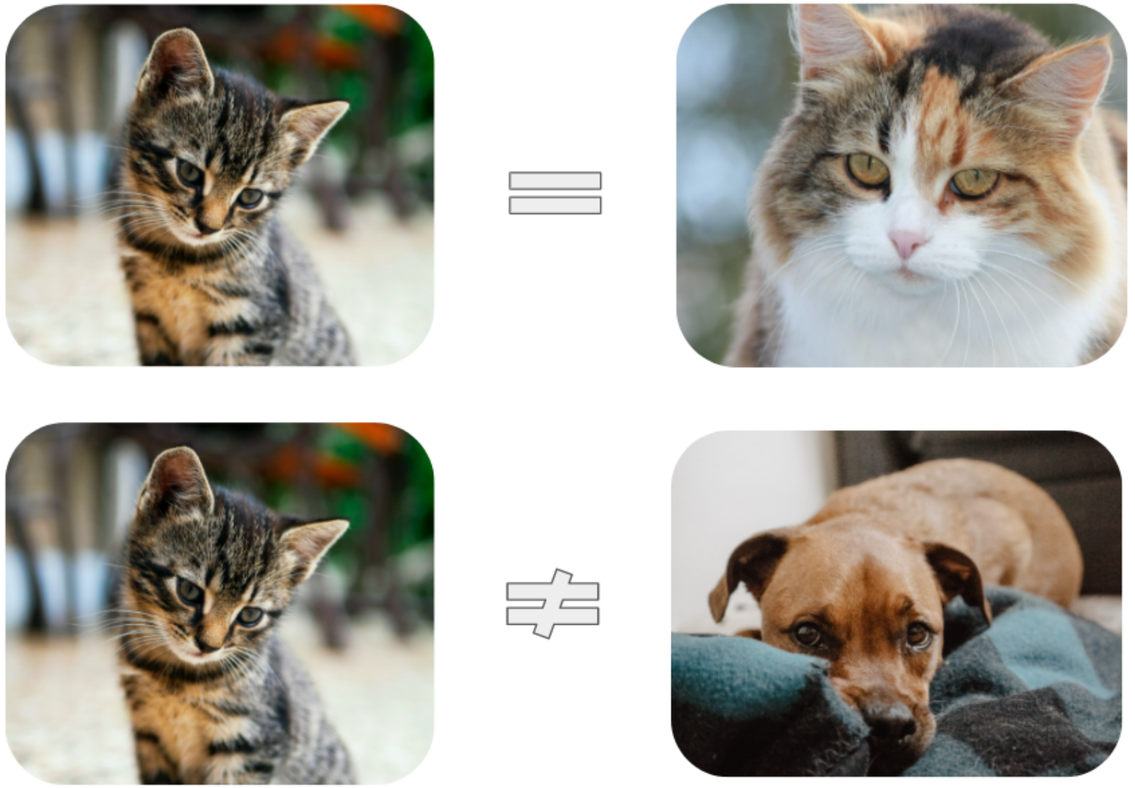
- Applicable to many faces without re-training
- Number of faces needed for face-related tasks —> SMALL

Eg: Face verification = thresholding, classification = nearest-neighbor


Vectors of similar faces
are near-by
Different faces
are far apart
Latent Space

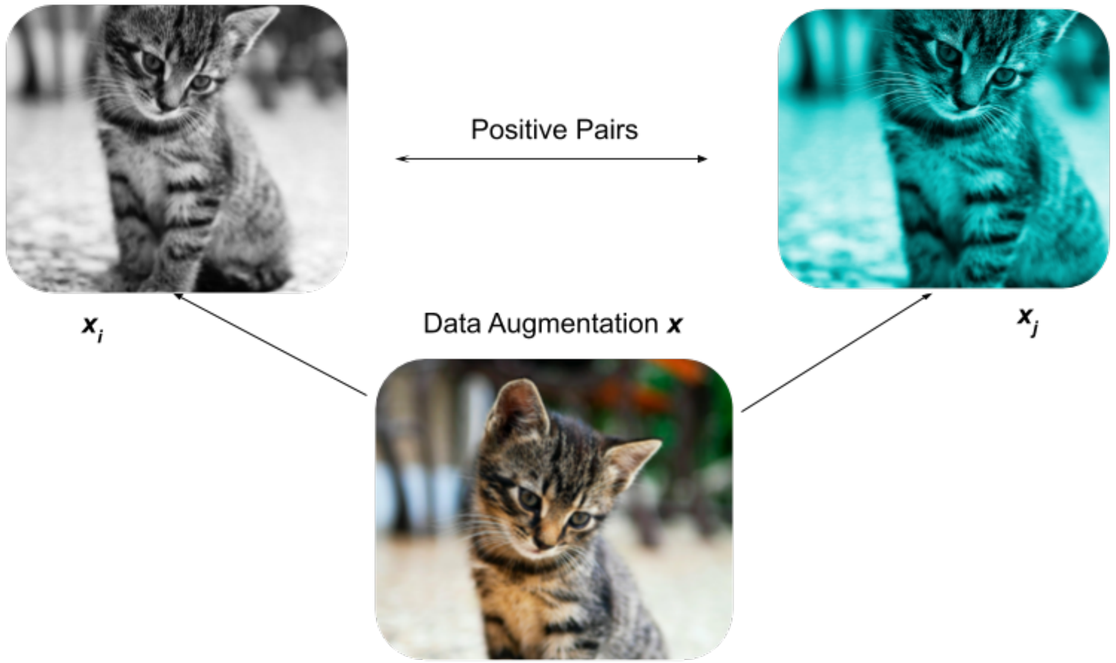
(Yash 2022)

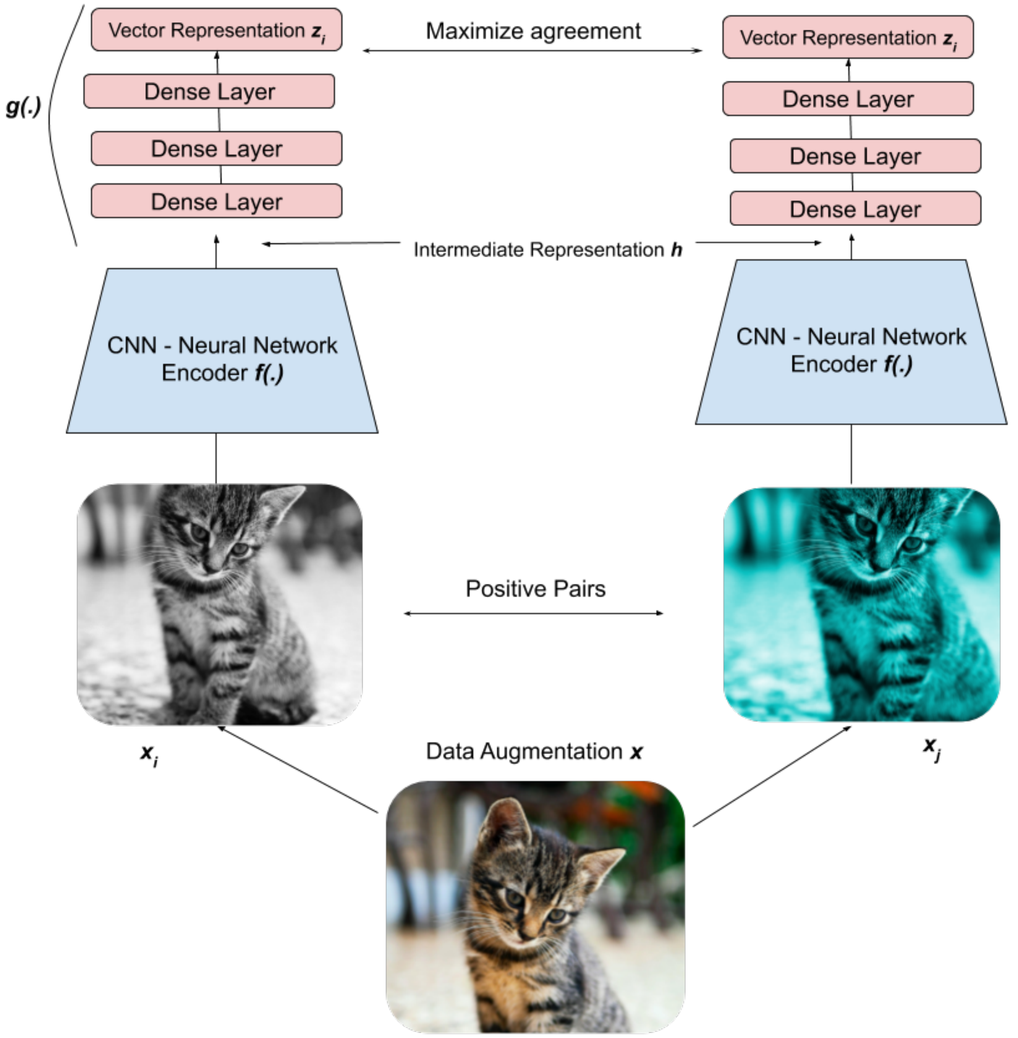
Humans are very good at self-supervised learning

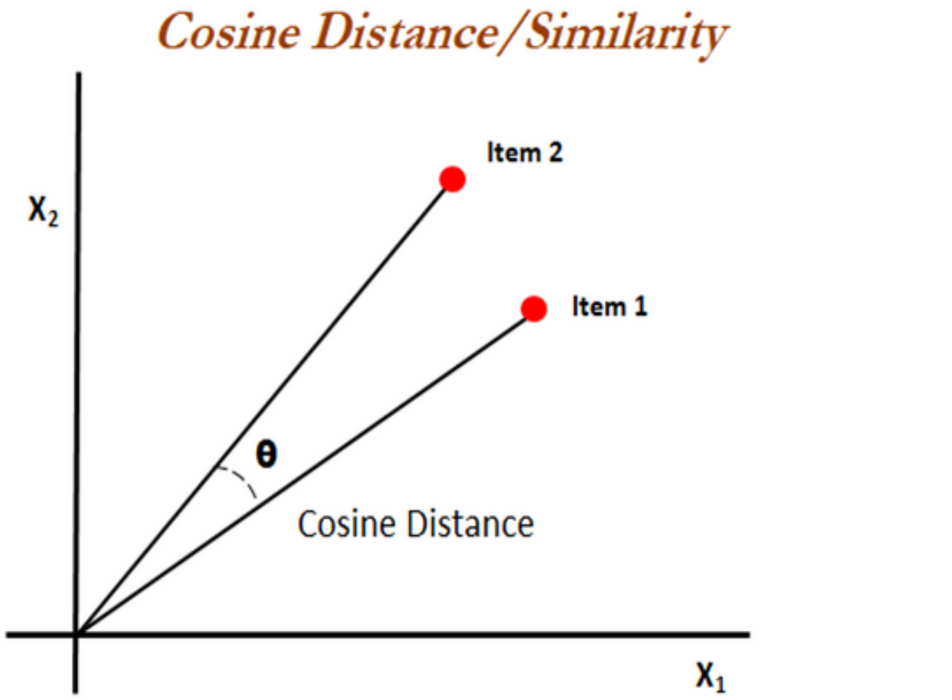


Contrastive learning in SSL

- SimCLR

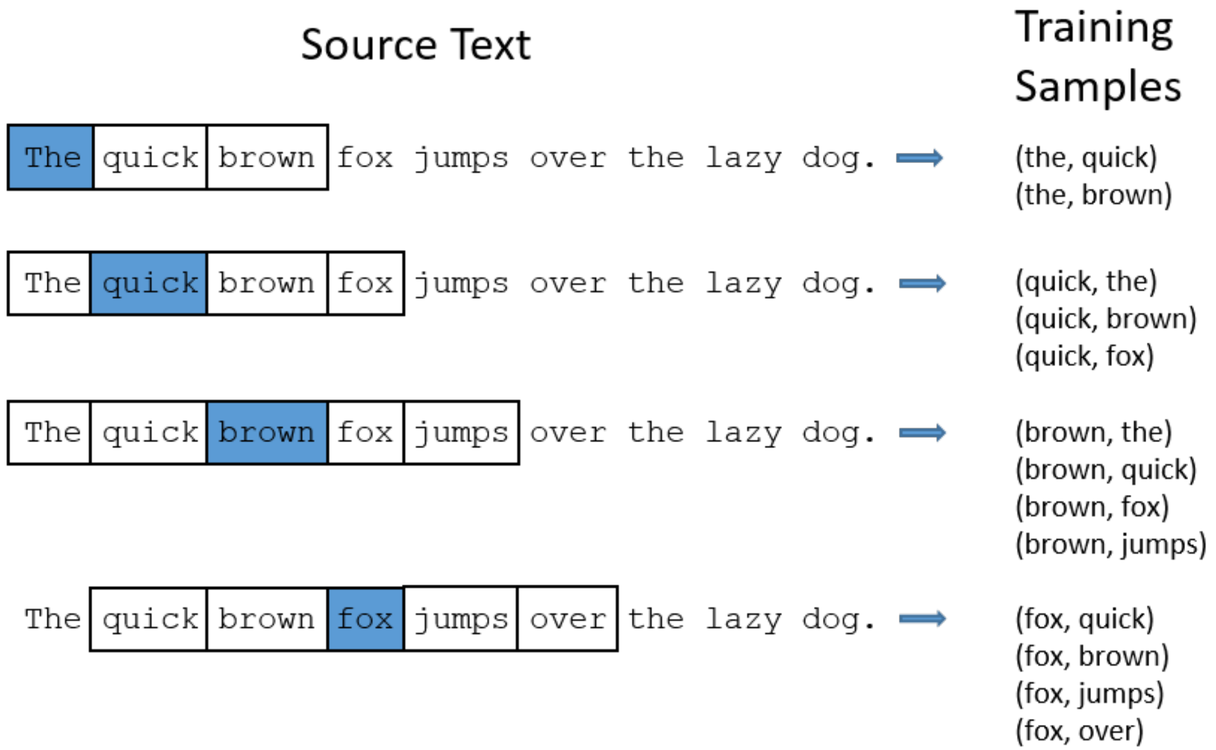






$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{||A|| ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Back to Word2vec



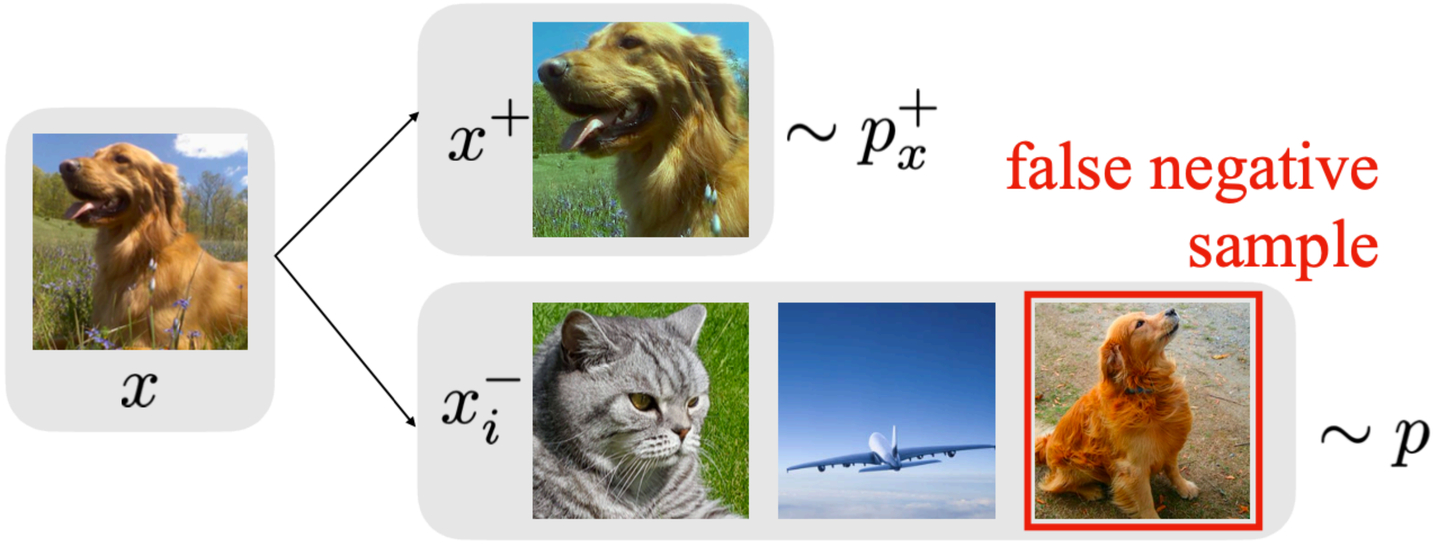
Careful with human bias

“Computer programmer" - “man” + “women” = “homemaker”

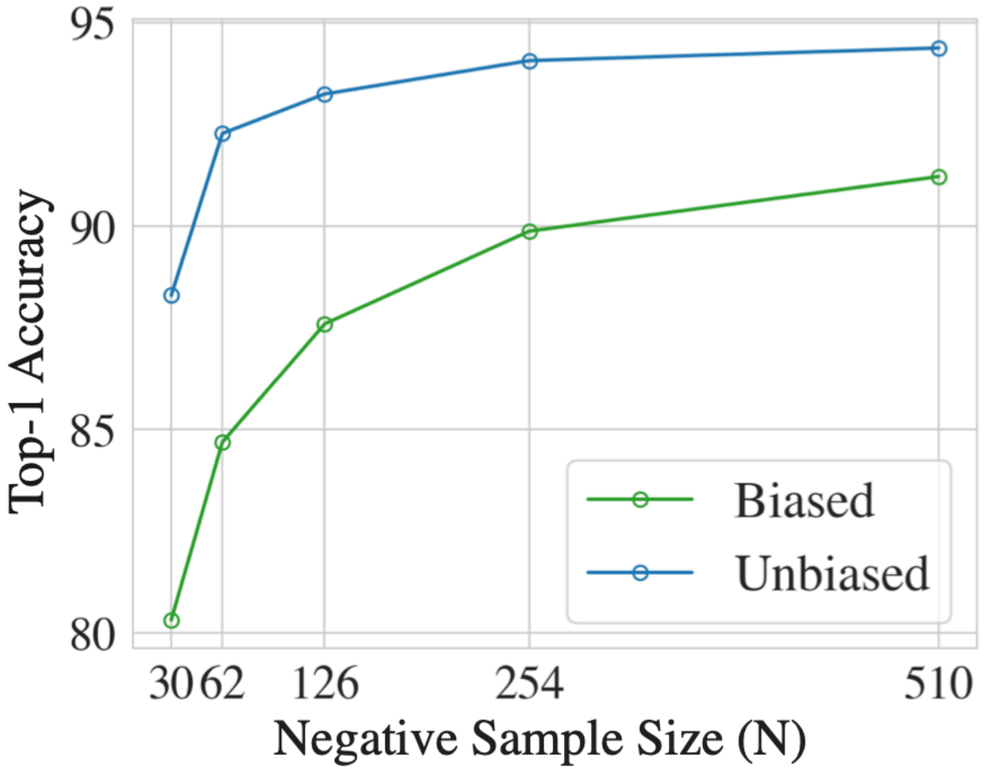
To go further

Sampling bias?

- Sampling negative pairs from the data can result in bias



(Chuang et al. NeurIPS 2020)



Example on CIFAR10

$$L^N_{\text{Unbiased}}(f) = \mathbb{E}_{\substack{x \sim p, x^+ \sim p^+_x \\ x^-_i \sim p^-_x}} \left[-\log \frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + \frac{Q}{N} \sum_{i=1}^{N} e^{f(x)^T f(x^-_i)}} \right]$$

$$L^N_{\text{Biased}}(f) \geq L^N_{\text{Unbiased}}(f) + \mathbb{E}_{x \sim p} \left[0 \wedge \log \frac{\mathbb{E}_{x^+ \sim p^+_x} \exp f(x)^\top f(x^+)}{\mathbb{E}_{x^- \sim p^-_x} \exp f(x)^\top f(x^-)} \right] - e^{3/2} \sqrt{\frac{\pi}{2N}}$$

Proposed a new loss to solve this issue

Paper available at https://arxiv.org/pdf/2007.00224.pdf

Cool papers

Towards Domain-Agnostic Contrastive Learning, (Verma et al. ICML 2021)

https://arxiv.org/pdf/2011.04419.pdf

Contrastive Learning as Goal-Conditioned Reinforcement Learning (Eysenbach et al. NeurIPS 2022)

https://arxiv.org/pdf/2206.07568.pdf