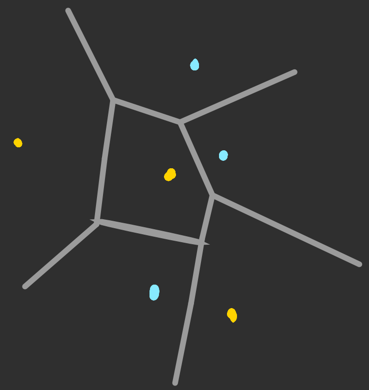
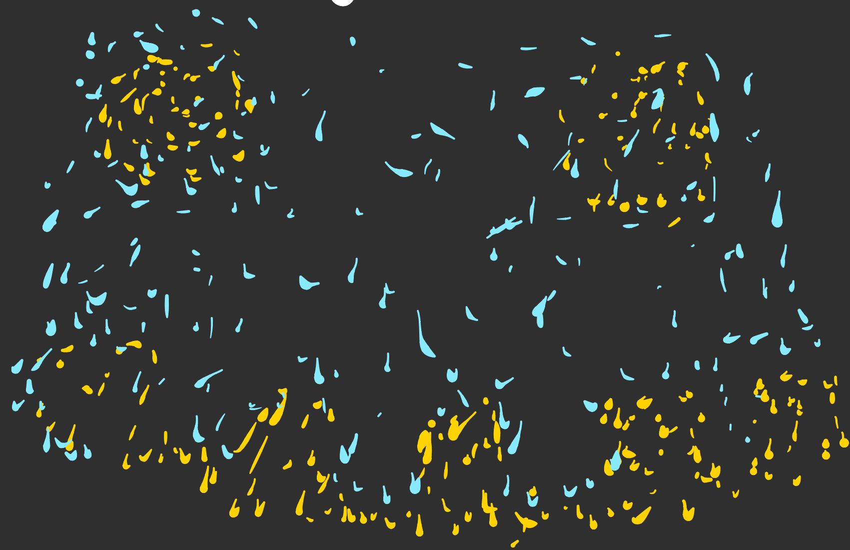
Nearest Neighbors



- ✓ Non-parametric
- ✓ Classification
- ✓ Regression

Key Idea

close-by ⇒ similar



Close in space ⇒ close in taste

The 1-nearest-neighbor Algorithm

Training

1. Store all (x_i, y_i) in dataset D.

... that's it.

Testing / Deployment

Input: $x_q \leftarrow$ "query"

1. Find closest x_i in training dataset.
2. Set $\hat{y} = \hat{f}(x_i)$

Output: $\hat{y} \leftarrow$ prediction

What does this look like?

The K-Nearest Neighbors Algorithm (regression)

Training

1. Store all (x_i, y_i) in dataset D.

... that's it.

Testing / Deployment

Input: $x_q \leftarrow$ "query", K

1. Find closest $(\tilde{x}_1, \ldots, \tilde{x}_K)$ in training dataset.
2. Set $\hat{y} = \frac{1}{K} \sum_{j=1}^{K} \hat{f}(\tilde{x}_j)$

Output: $\hat{y} \leftarrow$ prediction

The K-Nearest Neighbors Algorithm (classifier)

Training

1. Store all (x_i, y_i) in dataset D.

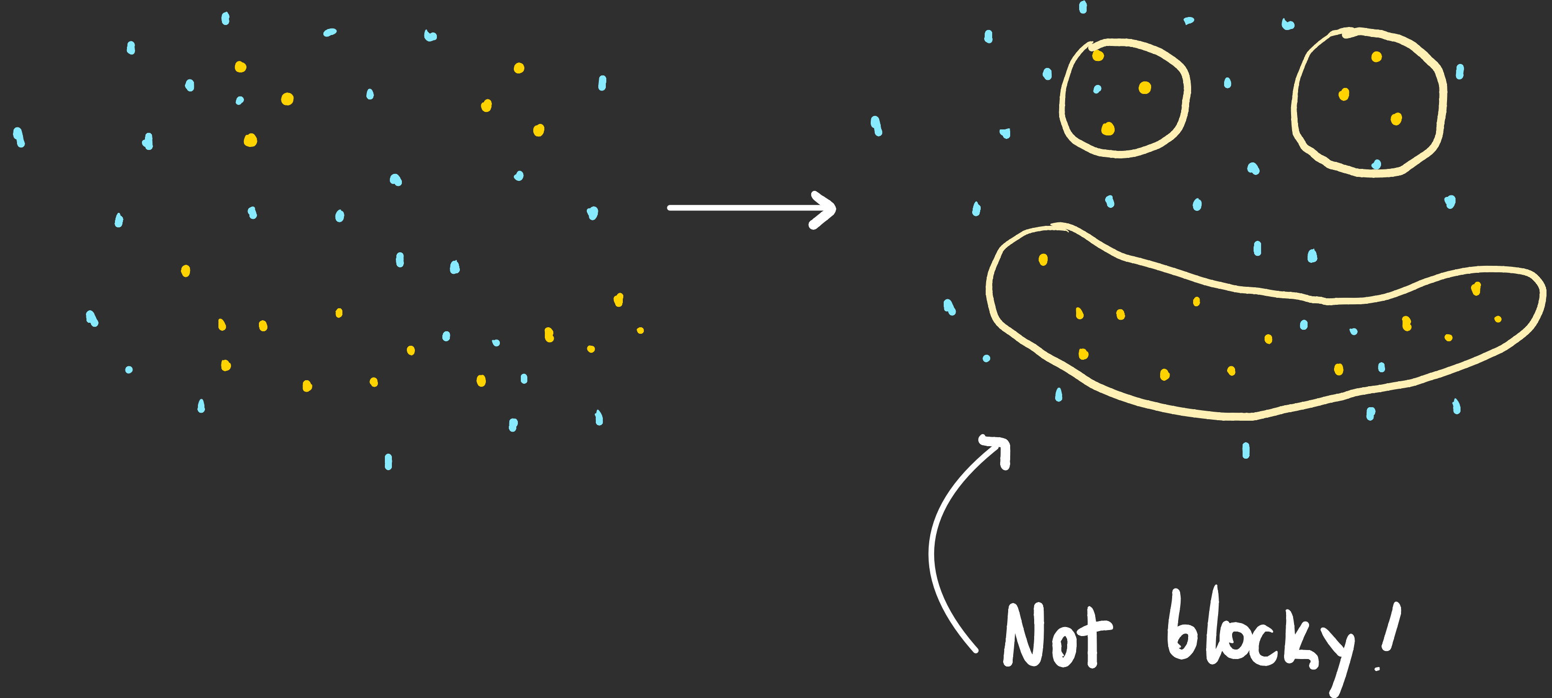
... that's it.

Testing / Deployment

Input: $x_q \leftarrow$ "query", K

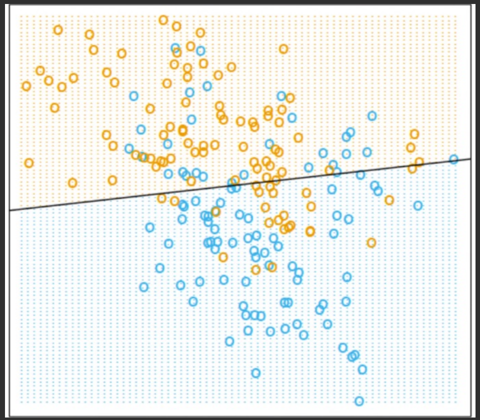
1. Find closest $(\tilde{x}_1, \ldots, \tilde{x}_K)$ in training dataset.
2. Set $\hat{y}$ = majority vote of $f(\tilde{x}_1), \ldots, f(\tilde{x}_K)$

Output: $\hat{y} \leftarrow$ prediction

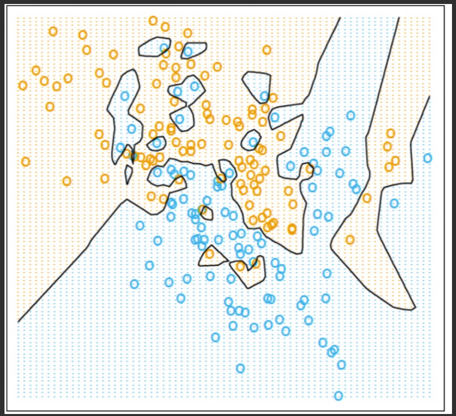

Not blocky!

Real Data

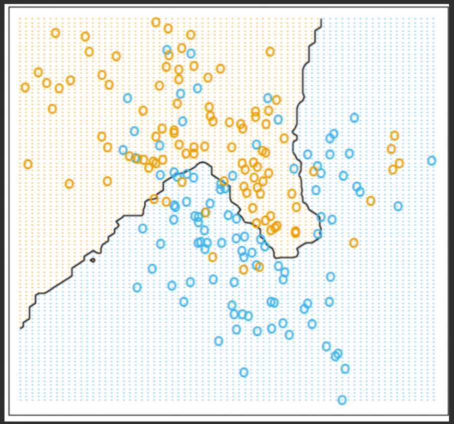
Linear Classifier



1-NN



15-NN



Diagrams from Hastie, Tibshirani & Friedman

What's the Decision Region?

$S \leftarrow$ Training set

Voronoi cell of $x \in S$ = all x' closer to x than any other in S.

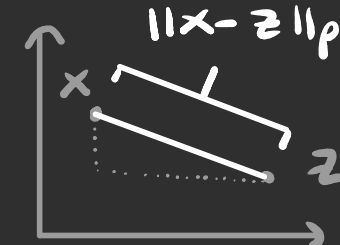
Region of class •: All Voronoi cells of •

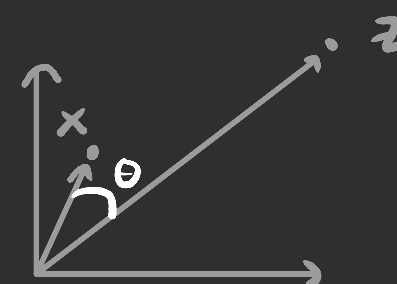
What is "closer"?

Any distance metric will work

E.g., if x, z are d-dimensional,

p-norm : $\|x - z\|_p = \left(\sum_{j=1}^{d} |x_j - z_j|^p \right)^{1/p}$





cosine similarity : $1 - \cos(\theta) = \frac{x^T z}{\|x\|_2 \|z\|_2}$

Behavior in the limit

$\varepsilon^*(x)$: average error of Bayes classifier

$\varepsilon_{NN}(x)$: error of NN

Nearest Neighbor Pattern Classification

T. M. COVER, MEMBER, IEEE, AND P. E. HART, MEMBER, IEEE

Thm 1:

$$\lim_{n \to \infty} \varepsilon_{NN}(x) \le 2\varepsilon^*(x)$$

Thm 2:

$$\lim_{\substack{n \to \infty \\ k \to \infty}} \varepsilon_{kNN}(x) = \varepsilon^*(x) \quad \text{if } \frac{k}{n} \to 0.$$

Advantages & Disadvantages

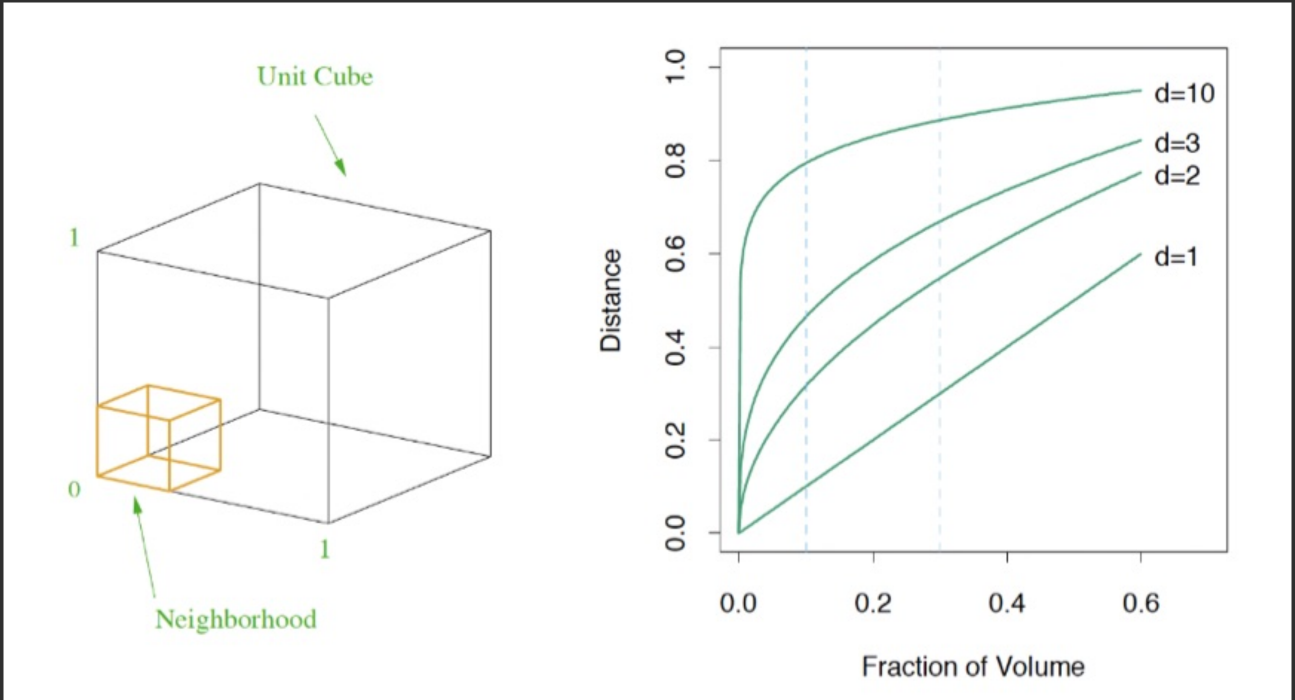
+ Fast training
+ Learns complex functions easily

– Slow at test time
– High storage cost
– Can perform poorly when x is high-dimensional

⚠ Requires a "good" distance metric!

Nowadays, learned!

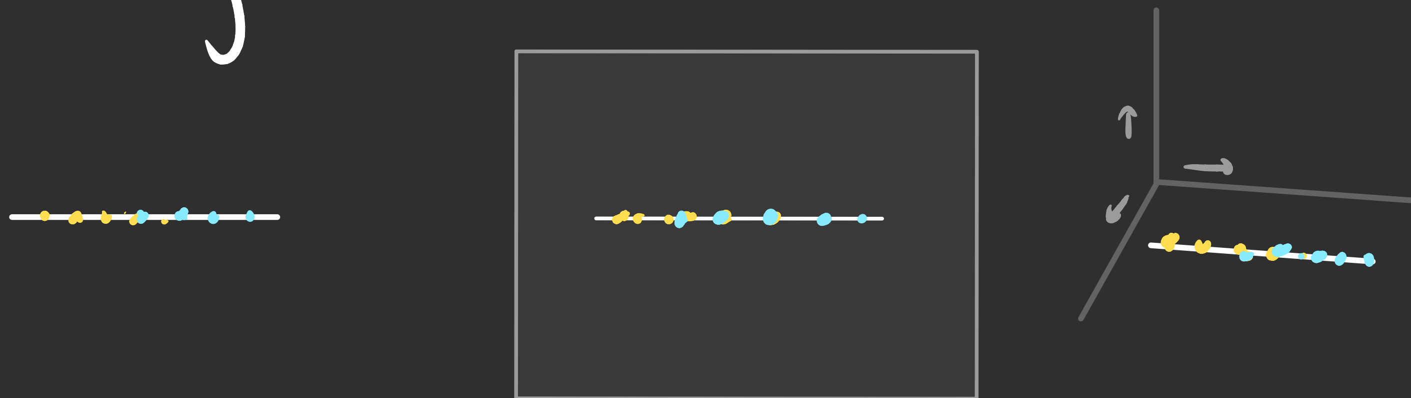
Curse of Dimensionality



Conclusion: distances are meaningless?

Good News!

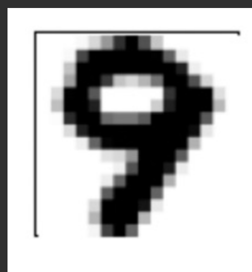
Only the "intrinsic dimension" matters

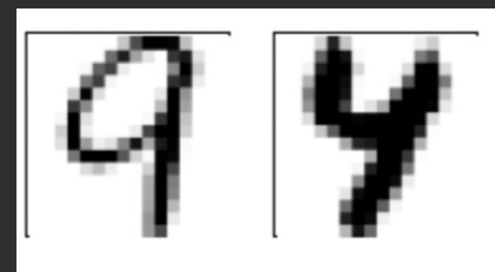


All intrinsically 1D

What about the "manifold of images?"

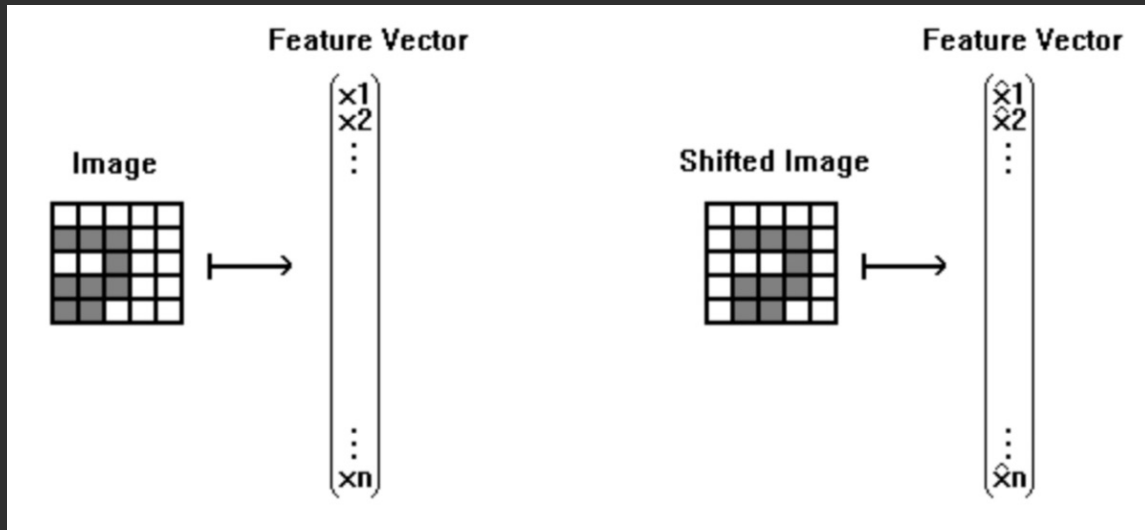
Euclidean not good!





Which are closer?

Why is this a problem?



Transformation Invariance

1 $\rightarrow$ 1

these should be close

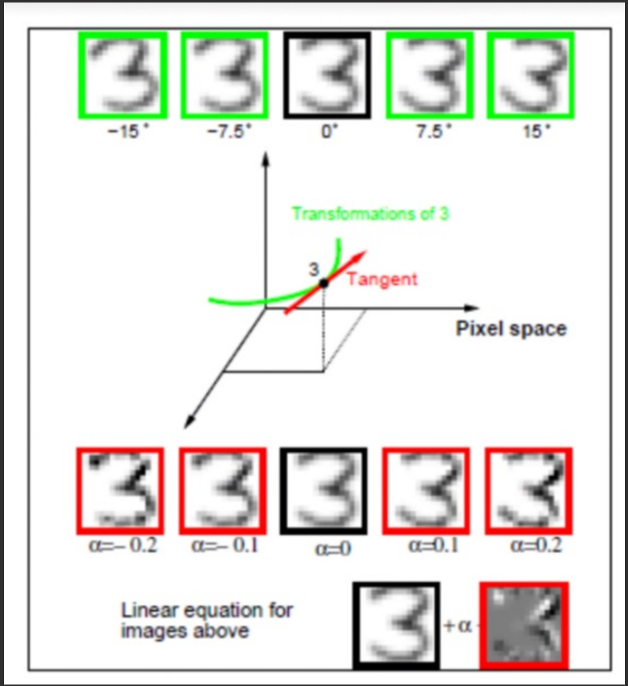
still far

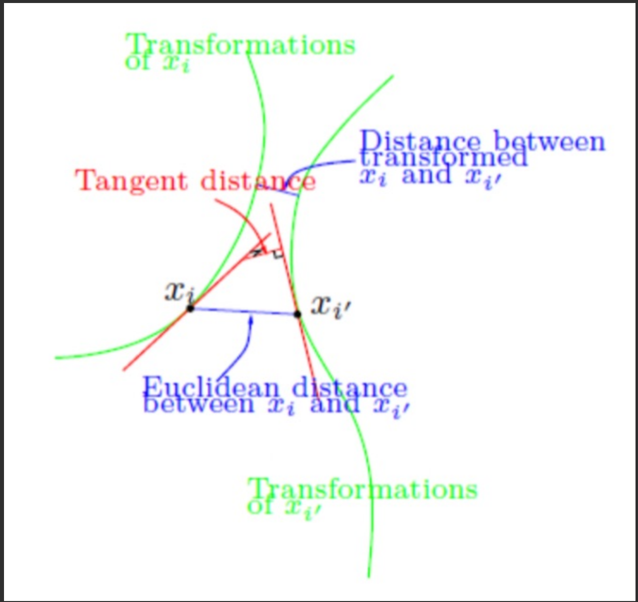
6 $\rightarrow$ 9

Building In Invariances

- Augment the dataset
- Build into features
  - e.g. ConvNets → shift invariance
- Design a good distance
  - e.g. "tangent distance"

Tangent Distance





Details in Simard et al. "Tangent Distance"

Tan distence results on digits

Feature	Classifier	Error Rate
Raw Pixels	SVM (linear)	11.3%
Raw Pixels	SVM (intersection)	8.7%
Raw Pixels	SVM (poly, $d = 3$) [7]	4.0%
Raw Pixels	VSV (poly, $d = 3$) [7]	3.2%
PHOG	SVM (linear)	3.4%
PHOG	SVM (intersection)	3.4%
PHOG	SVM (poly, $d = 5$)	3.2%
PHOG	SVM (rbf, $\gamma = 0.1$)	2.7%
Raw Pixels	Tangent Distance [23]*	2.6%
Raw Pixels	Boosted Neural Nets [8]*	2.6%
	Human Error Rate [3]	2.5%

Learned Embeddings

This CVPR2015 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

FaceNet: A Unified Embedding for Face Recognition and Clustering

Florian Schroff
fschroff@google.com
Google Inc.

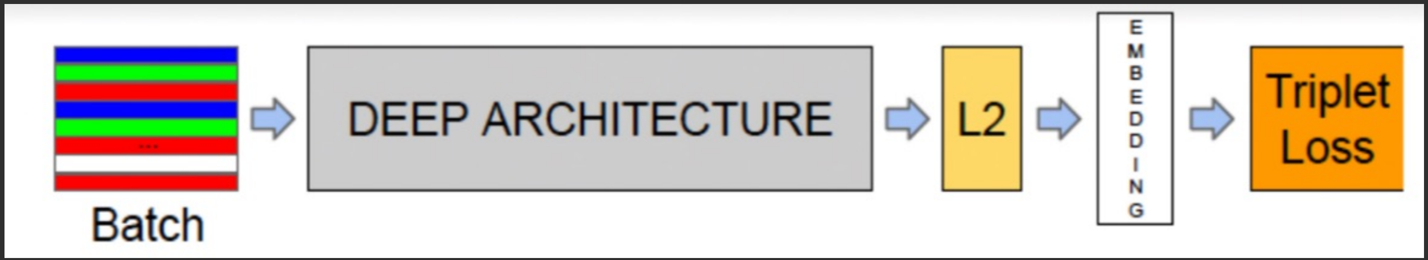
Dmitry Kalenichenko
dkalenichenko@google.com
Google Inc.

James Philbin
jphilbin@google.com
Google Inc.

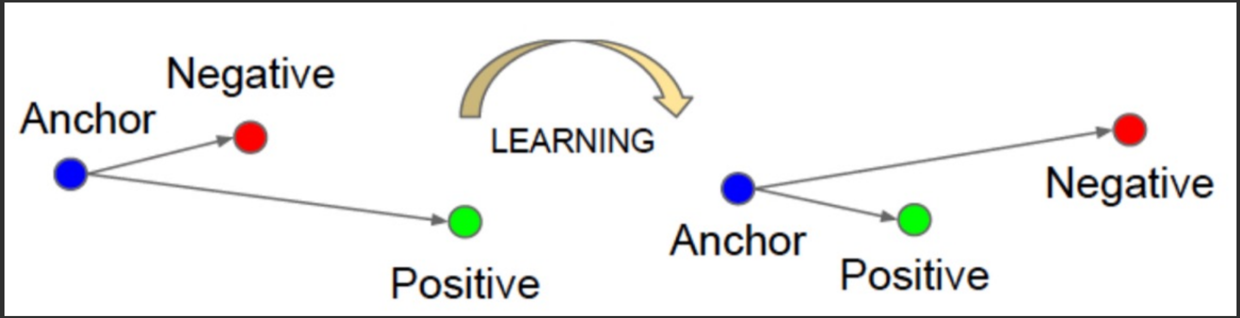

1.04
1.22
1.33
0.78

Contrastive Learning

embedding



trained with ...



Triplet loss:

$$\mathcal{L} = \sum_{i=1}^{N} \left(\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right)_+$$

Nowadays..

Masked Autoencoders Are Scalable Vision Learners

Kaiming He*,† Xinlei Chen* Saining Xie Yanghao Li Piotr Dollár Ross Girshick

*equal technical contribution †project lead

Facebook AI Research (FAIR)

A Simple Framework for Contrastive Learning of Visual Representations

Ting Chen [1] **Simon Kornblith** [1] **Mohammad Norouzi** [1] **Geoffrey Hinton** [1]

Siamese Neural Networks for One-shot Image Recognition

Gregory Koch GKOCH@CS.TORONTO.EDU
Richard Zemel ZEMEL@CS.TORONTO.EDU
Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Department of Computer Science, University of Toronto. Toronto, Ontario, Canada.