Title: Self-supervised learning

URL: https://en.wikipedia.org/wiki/Self-supervised_learning

PageID: 67902375

Categories: Category:Generative artificial intelligence, Category:Machine learning

-----

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning
Decision trees
Ensembles Bagging Boosting Random forest
Bagging
Boosting
Random forest
k -NN
Linear regression
Naive Bayes
Artificial neural networks
Logistic regression
Perceptron
Relevance vector machine (RVM)
Support vector machine (SVM)
BIRCH
CURE
Hierarchical
k -means
Fuzzy
Expectation–maximization (EM)
DBSCAN
OPTICS
Mean shift
Factor analysis
CCA
ICA
LDA
NMF
PCA
PGD
t-SNE
SDL
Graphical models Bayes net Conditional random field Hidden Markov
Bayes net
Conditional random field
Hidden Markov
RANSAC
k -NN

Local outlier factor

Isolation forest

Autoencoder

Deep learning

Feedforward neural network

Recurrent neural network LSTM GRU ESN reservoir computing

LSTM

GRU

ESN

reservoir computing

Boltzmann machine Restricted

Restricted

GAN

Diffusion model

SOM

Convolutional neural network U-Net LeNet AlexNet DeepDream

U-Net

LeNet

AlexNet

DeepDream

Neural field Neural radiance field Physics-informed neural networks

Neural radiance field

Physics-informed neural networks

Transformer Vision

Vision

Mamba

Spiking neural network

Memtransistor

Electrochemical RAM (ECRAM)

Q-learning

Policy gradient

SARSA

Temporal difference (TD)

Multi-agent Self-play

Self-play

Active learning

Crowdsourcing

Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

Self-supervised learning (SSL) is a paradigm in machine learning where a model is trained on a task using the data itself to generate supervisory signals, rather than relying on externally-provided labels. In the context of neural networks , self-supervised learning aims to leverage inherent structures or relationships within the input data to create meaningful training signals. SSL tasks are designed so that solving them requires capturing essential features or relationships in the data. The input data is typically augmented or transformed in a way that creates pairs of related samples, where one sample serves as the input, and the other is used to formulate the supervisory signal. This augmentation can involve introducing noise, cropping, rotation, or other transformations. Self-supervised learning more closely imitates the way humans learn to classify objects. [1]

During SSL, the model learns in two steps. First, the task is solved based on an auxiliary or pretext classification task using pseudo-labels, which help to initialize the model parameters . [2] [3] Next, the actual task is performed with supervised or unsupervised learning . [4] [5] [6]

Self-supervised learning has produced promising results in recent years, and has found practical application in fields such as audio processing , and is being used by Facebook and others for speech recognition . [7]

Types

Autoassociative self-supervised learning

Autoassociative self-supervised learning is a specific category of self-supervised learning where a neural network is trained to reproduce or reconstruct its own input data. [8] In other words, the model is tasked with learning a representation of the data that captures its essential features or structure, allowing it to regenerate the original input.

The term "autoassociative" comes from the fact that the model is essentially associating the input data with itself. This is often achieved using autoencoders , which are a type of neural network architecture used for representation learning. Autoencoders consist of an encoder network that maps the input data to a lower-dimensional representation (latent space), and a decoder network that reconstructs the input from this representation.

The training process involves presenting the model with input data and requiring it to reconstruct the same data as closely as possible. The loss function used during training typically penalizes the difference between the original input and the reconstructed output (e.g. mean squared error). By minimizing this reconstruction error, the autoencoder learns a meaningful representation of the data in its latent space .

Contrastive self-supervised learning

For a binary classification task, training data can be divided into positive examples and negative examples. Positive examples are those that match the target. For example, if training a classifier to identify birds, the positive training data would include images that contain birds. Negative examples would be images that do not. [9] Contrastive self-supervised learning uses both positive and negative examples. The loss function in contrastive learning is used to minimize the distance between positive sample pairs, while maximizing the distance between negative sample pairs. [9]

An early example uses a pair of 1-dimensional convolutional neural networks to process a pair of images and maximize their agreement. [10]

Contrastive Language-Image Pre-training (CLIP) allows joint pretraining of a text encoder and an image encoder, such that a matching image-text pair have image encoding vector and text encoding vector that span a small angle (having a large cosine similarity).

InfoNCE (Noise-Contrastive Estimation) [11] is a method to optimize two models jointly, based on Noise Contrastive Estimation (NCE). [12] Given a set X = { x 1 , … x N } {\displaystyle X=\left\{x_{1},\ldots x_{N}\right\}} of N {\displaystyle N} random samples containing one positive sample from p (x t + k ■ c t) {\displaystyle p\left(x_{t+k}\mid c_{t}\right)} and N − 1 {\displaystyle N-1} negative samples from the 'proposal' distribution p (x t + k) {\displaystyle p\left(x_{t+k}\right)} , it minimizes the following loss function:

L N = − E X [log ■ f k (x t + k , c t) ∑ x j ∈ X f k (x j , c t)] {\displaystyle {\mathcal {L}}_{\mathrm {N} }=-\mathbb {E} _{X}\left[\log {\frac {f_{k}\left(x_{t+k},c_{t}\right)}{\sum _{x_{j}\in X}f_{k}\left(x_{j},c_{t}\right)}}\right]}

Non-contrastive self-supervised learning

Non-contrastive self-supervised learning (NCSSL) uses only positive examples. Counterintuitively, NCSSL converges on a useful local minimum rather than reaching a trivial solution, with zero loss. For the example of binary classification, it would trivially learn to classify each example as positive. Effective NCSSL requires an extra predictor on the online side that does not back-propagate on the target side. [9]

Comparison with other forms of machine learning

SSL belongs to supervised learning methods insofar as the goal is to generate a classified output from the input. At the same time, however, it does not require the explicit use of labeled input-output pairs. Instead, correlations, metadata embedded in the data, or domain knowledge present in the input are implicitly and autonomously extracted from the data. These supervisory signals, extracted from the data, can then be used for training. [1]

SSL is similar to unsupervised learning in that it does not require labels in the sample data. Unlike unsupervised learning, however, learning is not done using inherent data structures.

Semi-supervised learning combines supervised and unsupervised learning, requiring only a small portion of the learning data be labeled . [3]

In transfer learning , a model designed for one task is reused on a different task. [13]

Training an autoencoder intrinsically constitutes a self-supervised process, because the output pattern needs to become an optimal reconstruction of the input pattern itself. However, in current jargon, the term 'self-supervised' often refers to tasks based on a pretext-task training setup. This involves the (human) design of such pretext task(s), unlike

the case of fully self-contained autoencoder training. [8]

In reinforcement learning , self-supervising learning from a combination of losses can create abstract representations where only the most important information about the state are kept in a compressed way. [14]

Examples

Self-supervised learning is particularly suitable for speech recognition. For example, Facebook developed wav2vec , a self-supervised algorithm, to perform speech recognition using two deep convolutional neural networks that build on each other. [7]

Google 's Bidirectional Encoder Representations from Transformers (BERT) model is used to better understand the context of search queries. [15]

OpenAI 's GPT-3 is an autoregressive language model that can be used in language processing. It can be used to translate texts or answer questions, among other things. [16]

Bootstrap Your Own Latent (BYOL) is a NCSSL that produced excellent results on ImageNet and on transfer and semi-supervised benchmarks. [17]

The Yarowsky algorithm is an example of self-supervised learning in natural language processing . From a small number of labeled examples, it learns to predict which word sense of a polysemous word is being used at a given point in text.

DirectPred is a NCSSL that directly sets the predictor weights instead of learning it via typical gradient descent . [9]

Self-GenomeNet is an example of self-supervised learning in genomics. [18]

Self-supervised learning continues to gain prominence as a new approach across diverse fields. Its ability to leverage unlabeled data effectively opens new possibilities for advancement in machine learning, especially in data-driven application domains.

References

Further reading

Balestriero, Randall; Ibrahim, Mark; Sobal, Vlad; Morcos, Ari; Shekhar, Shashank; Goldstein, Tom; Bordes, Florian; Bardes, Adrien; Mialon, Gregoire; Tian, Yuandong; Schwarzschild, Avi; Wilson, Andrew Gordon; Geiping, Jonas; Garrido, Quentin; Fernandez, Pierre (24 April 2023). "A Cookbook of Self-Supervised Learning". arXiv : 2304.12210 [cs.LG].

External links

Doersch, Carl; Zisserman, Andrew (October 2017). "Multi-task Self-Supervised Visual Learning". 2017 IEEE International Conference on Computer Vision (ICCV) . pp. 2070– 2079. arXiv : 1708.07860 . doi : 10.1109/ICCV.2017.226 . ISBN 978-1-5386-1032-9 . S2CID 473729 .

Doersch, Carl; Gupta, Abhinav; Efros, Alexei A. (December 2015). "Unsupervised Visual Representation Learning by Context Prediction". 2015 IEEE International Conference on Computer Vision (ICCV) . pp. 1422– 1430. arXiv : 1505.05192 . doi : 10.1109/ICCV.2015.167 . ISBN 978-1-4673-8391-2 . S2CID 9062671 .

Zheng, Xin; Wang, Yong; Wang, Guoyou; Liu, Jianguo (1 April 2018). "Fast and robust segmentation of white blood cell images by self-supervised learning" . Micron . 107 : 55– 71. doi : 10.1016/j.micron.2018.01.010 . ISSN 0968-4328 . PMID 29425969 . S2CID 3796689 .

Yarowsky, David (1995). "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods" . Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics . Cambridge, MA: Association for Computational Linguistics: 189– 196. doi : 10.3115/981658.981684 . Retrieved 1 November 2022 .

v

t

e

History timeline

timeline

Companies

Projects

Parameter Hyperparameter

Hyperparameter

Loss functions

Regression Bias–variance tradeoff Double descent Overfitting

Bias–variance tradeoff

Double descent

Overfitting

Clustering

Gradient descent SGD Quasi-Newton method Conjugate gradient method

SGD

Quasi-Newton method

Conjugate gradient method

Backpropagation

Attention

Convolution

Normalization Batchnorm

Batchnorm

Activation Softmax Sigmoid Rectifier

Softmax

Sigmoid

Rectifier

Gating

Weight initialization

Regularization

Datasets Augmentation

Augmentation

Prompt engineering

Reinforcement learning Q-learning SARSA Imitation Policy gradient

Q-learning

SARSA

Imitation

Policy gradient

Diffusion

Latent diffusion model

Autoregression

Adversary

RAG

Uncanny valley

RLHF

Self-supervised learning

Reflection

Recursive self-improvement

Hallucination

Word embedding

Vibe coding

Machine learning In-context learning

In-context learning

Artificial neural network Deep learning

Deep learning

Language model Large language model NMT

Large language model

NMT

Reasoning language model

Model Context Protocol

Intelligent agent

Artificial human companion

Humanity's Last Exam

Artificial general intelligence (AGI)

AlexNet

WaveNet

Human image synthesis

HWR

OCR

Computer vision

Speech synthesis 15.ai ElevenLabs

15.ai

ElevenLabs

Speech recognition Whisper

Whisper

Facial recognition

AlphaFold

Text-to-image models Aurora DALL-E Firefly Flux Ideogram Imagen Midjourney Recraft Stable Diffusion

Aurora

DALL-E

Firefly

Flux

Ideogram

Imagen

Midjourney

Recraft

Stable Diffusion

Text-to-video models Dream Machine Runway Gen Hailuo AI Kling Sora Veo

Dream Machine

Runway Gen

Hailuo AI

Kling

Sora

Veo

Music generation Riffusion Suno AI Udio

Riffusion

Suno AI

Udio

Word2vec

Seq2seq

GloVe

BERT

T5

Llama

Chinchilla AI

PaLM

GPT 1 2 3 J ChatGPT 4 4o o1 o3 4.5 4.1 o4-mini 5

1

2

3

J

ChatGPT

4

4o

o1

o3

4.5

4.1

o4-mini

5

Claude

Gemini Gemini (language model) Gemma

Gemini (language model)

Gemma

Grok

LaMDA

BLOOM

DBRX

Project Debater

IBM Watson

IBM Watsonx

Granite

PanGu-Σ

DeepSeek

Qwen

AlphaGo

AlphaZero

OpenAI Five

Self-driving car

MuZero
Action selection AutoGPT
AutoGPT
Robot control
Alan Turing
Warren Sturgis McCulloch
Walter Pitts
John von Neumann
Claude Shannon
Shun'ichi Amari
Kunihiko Fukushima
Takeo Kanade
Marvin Minsky
John McCarthy
Nathaniel Rochester
Allen Newell
Cliff Shaw
Herbert A. Simon
Oliver Selfridge
Frank Rosenblatt
Bernard Widrow
Joseph Weizenbaum
Seymour Papert
Seppo Linnainmaa
Paul Werbos
Geoffrey Hinton
John Hopfield
Jürgen Schmidhuber
Yann LeCun
Yoshua Bengio
Lotfi A. Zadeh
Stephen Grossberg
Alex Graves
James Goodnight
Andrew Ng
Fei-Fei Li
Alex Krizhevsky
Ilya Sutskever

Oriol Vinyals
Quoc V. Le
Ian Goodfellow
Demis Hassabis
David Silver
Andrej Karpathy
Ashish Vaswani
Noam Shazeer
Aidan Gomez
John Schulman
Mustafa Suleyman
Jan Leike
Daniel Kokotajlo
François Chollet
Neural Turing machine
Differentiable neural computer
Transformer Vision transformer (ViT)
Vision transformer (ViT)
Recurrent neural network (RNN)
Long short-term memory (LSTM)
Gated recurrent unit (GRU)
Echo state network
Multilayer perceptron (MLP)
Convolutional neural network (CNN)
Residual neural network (RNN)
Highway network
Mamba
Autoencoder
Variational autoencoder (VAE)
Generative adversarial network (GAN)
Graph neural network (GNN)
Category