Title: Learning curve (machine learning)

URL: https://en.wikipedia.org/wiki/Learning_curve_(machine_learning)

PageID: 59968610

Categories: Category:Machine learning, Category:Model selection

-----

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

In machine learning (ML), a learning curve (or training curve) is a graphical representation that shows how a model's performance on a training set (and usually a validation set) changes with the number of training iterations (epochs) or the amount of training data. [1] Typically, the number of training epochs or training set size is plotted on the x -axis, and the value of the loss function (and possibly some other metric such as the cross-validation score) on the y -axis.

Synonyms include error curve , experience curve , improvement curve and generalization curve . [2]

More abstractly, learning curves plot the difference between learning effort and predictive performance, where "learning effort" usually means the number of training samples, and "predictive performance" means accuracy on testing samples. [3]

Learning curves have many useful purposes in ML, including: [4] [5] [6]

choosing model parameters during design,

adjusting optimization to improve convergence,

and diagnosing problems such as overfitting (or underfitting).

Learning curves can also be tools for determining how much a model benefits from adding more training data, and whether the model suffers more from a variance error or a bias error . If both the validation score and the training score converge to a certain value, then the model will no longer significantly benefit from more training data. [7]

Formal definition

When creating a function to approximate the distribution of some data, it is necessary to define a loss function L (f θ (X) , Y) {\displaystyle L(f_{\theta }(X),Y)} to measure how good the model output is (e.g., accuracy for classification tasks or mean squared error for regression). We then define an optimization process which finds model parameters θ {\displaystyle \theta } such that L (f θ (X) , Y) {\displaystyle L(f_{\theta }(X),Y)} is minimized, referred to as θ ∗ {\displaystyle \theta ^{*}} .

Training curve for amount of data

If the training data is

{ x 1 , x 2 , … , x n } , { y 1 , y 2 , … y n } {\displaystyle \{x_{1},x_{2},\dots ,x_{n}\},\{y_{1},y_{2},\dots y_{n}\}}

and the validation data is

{ x 1 ′ , x 2 ′ , … x m ′ } , { y 1 ′ , y 2 ′ , … y m ′ } {\displaystyle \{x_{1}',x_{2}',\dots x_{m}'\},\{y_{1}',y_{2}',\dots y_{m}'\}} ,

a learning curve is the plot of the two curves

i ■ L (f θ ∗ (X i , Y i) (X i) , Y i) {\displaystyle i\mapsto L(f_{\theta ^{*}(X_{i},Y_{i})}(X_{i}),Y_{i})}

i ■ L (f θ ∗ (X i , Y i) (X i ′) , Y i ′) {\displaystyle i\mapsto L(f_{\theta ^{*}(X_{i},Y_{i})}(X_{i}'),Y_{i}')}

where X i = { x 1 , x 2 , … x i } {\displaystyle X_{i}=\{x_{1},x_{2},\dots x_{i}\}}

Training curve for number of iterations

Many optimization algorithms are iterative, repeating the same step (such as backpropagation) until the process converges to an optimal value. Gradient descent is one such algorithm. If θ i ∗ {\displaystyle \theta _{i}^{*}} is the approximation of the optimal θ {\displaystyle \theta } after i {\displaystyle i} steps, a learning curve is the plot of

i ■ L (f θ i ∗ (X , Y) (X) , Y) {\displaystyle i\mapsto L(f_{\theta _{i}^{*}(X,Y)}(X),Y)}

i ■ L (f θ i ∗ (X , Y) (X ′) , Y ′) {\displaystyle i\mapsto L(f_{\theta _{i}^{*}(X,Y)}(X'),Y')}

See also

Overfitting

Bias–variance tradeoff

Model selection

Cross-validation (statistics)

Validity (statistics)

Verification and validation

Double descent

References