Title: Restricted Boltzmann machine

URL: https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine

PageID: 33742232

Categories: Category:Neural network architectures, Category:Stochastic models, Category:Supervised learning, Category:Unsupervised 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

A restricted Boltzmann machine (RBM) (also called a restricted Sherrington–Kirkpatrick model with external field or restricted stochastic Ising–Lenz–Little model) is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputs. [1]

RBMs were initially proposed under the name Harmonium by Paul Smolensky in 1986, [2] and rose to prominence after Geoffrey Hinton and collaborators used fast learning algorithms for them in the mid-2000s. RBMs have found applications in dimensionality reduction , [3] classification , [4] collaborative filtering , [5] feature learning , [6] topic modelling , [7] immunology , [8] and even many■body quantum mechanics . [9] [10] [11]

They can be trained in either supervised or unsupervised ways, depending on the task. [citation needed]

As their name implies, RBMs are a variant of Boltzmann machines , with the restriction that their neurons must form a bipartite graph :

a pair of nodes from each of the two groups of units (commonly referred to as the "visible" and "hidden" units respectively) may have a symmetric connection between them; and

there are no connections between nodes within a group.

By contrast, "unrestricted" Boltzmann machines may have connections between hidden units . This restriction allows for more efficient training algorithms than are available for the general class of Boltzmann machines, in particular the gradient-based contrastive divergence algorithm. [12]

Restricted Boltzmann machines can also be used in deep learning networks. In particular, deep belief networks can be formed by "stacking" RBMs and optionally fine-tuning the resulting deep network with gradient descent and backpropagation . [13]

Structure

The standard type of RBM has binary-valued (Boolean) hidden and visible units, and consists of a matrix of weights W {\displaystyle W} of size m × n {\displaystyle m\times n} . Each weight element (w i , j) {\displaystyle (w_{i,j})} of the matrix is associated with the connection between the visible (input) unit v i {\displaystyle v_{i}} and the hidden unit h j {\displaystyle h_{j}} . In addition, there are bias weights (offsets) a i {\displaystyle a_{i}} for v i {\displaystyle v_{i}} and b j {\displaystyle b_{j}} for h j {\displaystyle h_{j}} . Given the weights and biases, the energy of a configuration (pair of Boolean vectors) (v , h) is defined as

or, in matrix notation,

This energy function is analogous to that of a Hopfield network . As with general Boltzmann machines, the joint probability distribution for the visible and hidden vectors is defined in terms of the energy function as follows, [14]

where Z {\displaystyle Z} is a partition function defined as the sum of e – E (v , h) {\displaystyle e^{-E(v,h)}} over all possible configurations, which can be interpreted as a normalizing constant to ensure that the probabilities sum to 1. The marginal probability of a visible vector is the sum of P (v , h) {\displaystyle P(v,h)} over all possible hidden layer configurations, [14]

and vice versa. Since the underlying graph structure of the RBM is bipartite (meaning there are no intra-layer connections), the hidden unit activations are mutually independent given the visible unit activations. Conversely, the visible unit activations are mutually independent given the hidden unit activations. [12] That is, for m visible units and n hidden units, the conditional probability of a configuration of the visible units v , given a configuration of the hidden units h , is

Conversely, the conditional probability of h given v is

The individual activation probabilities are given by

where σ {\displaystyle \sigma } denotes the logistic sigmoid .

The visible units of Restricted Boltzmann Machine can be multinomial , although the hidden units are Bernoulli . [clarification needed] In this case, the logistic function for visible units is replaced by the softmax function

where K is the number of discrete values that the visible values have. They are applied in topic modeling, [7] and recommender systems . [5]

Relation to other models

Restricted Boltzmann machines are a special case of Boltzmann machines and Markov random fields . [15] [16]

The graphical model of RBMs corresponds to that of factor analysis . [17]

Training algorithm

Restricted Boltzmann machines are trained to maximize the product of probabilities assigned to some training set V {\displaystyle V} (a matrix, each row of which is treated as a visible vector v {\displaystyle v}),

or equivalently, to maximize the expected log probability of a training sample v {\displaystyle v} selected randomly from V {\displaystyle V} : [15] [16]

The algorithm most often used to train RBMs, that is, to optimize the weight matrix W {\displaystyle W} , is the contrastive divergence (CD) algorithm due to Hinton , originally developed to train PoE (product of experts) models. [18] [19] The algorithm performs Gibbs sampling and is used inside a gradient descent procedure (similar to the way backpropagation is used inside such a procedure when training feedforward neural nets) to compute weight update.

The basic, single-step contrastive divergence (CD-1) procedure for a single sample can be summarized as follows:

Take a training sample v , compute the probabilities of the hidden units and sample a hidden activation vector h from this probability distribution.

Compute the outer product of v and h and call this the positive gradient .

From h , sample a reconstruction v' of the visible units, then resample the hidden activations h' from this. (Gibbs sampling step)

Compute the outer product of v' and h' and call this the negative gradient .

Let the update to the weight matrix W {\displaystyle W} be the positive gradient minus the negative gradient, times some learning rate: Δ W = ■ (v h T − v ′ h ′ T) {\displaystyle \Delta W=\epsilon (vh^{\mathsf {T}}-v'h'^{\mathsf {T}})} .

Update the biases a and b analogously: Δ a = ■ (v − v ′) {\displaystyle \Delta a=\epsilon (v-v')} , Δ b = ■ (h − h ′) {\displaystyle \Delta b=\epsilon (h-h')} .

A Practical Guide to Training RBMs written by Hinton can be found on his homepage. [14]

Stacked Restricted Boltzmann Machine

The difference between the Stacked Restricted Boltzmann Machines and RBM is that RBM has lateral connections within a layer that are prohibited to make analysis tractable. On the other hand, the Stacked Boltzmann consists of a combination of an unsupervised three-layer network with symmetric weights and a supervised fine-tuned top layer for recognizing three classes.

The usage of Stacked Boltzmann is to understand Natural languages , retrieve documents , image generation, and classification. These functions are trained with unsupervised pre-training and/or supervised fine-tuning. Unlike the undirected symmetric top layer, with a two-way unsymmetric layer for connection for RBM. The restricted Boltzmann's connection is three-layers with asymmetric weights, and two networks are combined into one.

Stacked Boltzmann does share similarities with RBM, the neuron for Stacked Boltzmann is a stochastic binary Hopfield neuron, which is the same as the Restricted Boltzmann Machine. The energy from both Restricted Boltzmann and RBM is given by Gibb's probability measure: E = − 1 2 ∑ i , j w i j s i s j + ∑ i θ i s i {\displaystyle E=-{\frac {1}{2}}\sum _{i,j}{w_{ij}{s_{i}}{s_{j}}}+\sum _{i}{\theta _{i}}{s_{i}}} . The training process of Restricted Boltzmann is similar to RBM. Restricted Boltzmann train one layer at a time and approximate equilibrium state with a 3-segment pass, not performing back propagation. Restricted Boltzmann uses both supervised and unsupervised on different RBM for pre-training for classification and recognition. The training uses contrastive divergence with Gibbs sampling: Δw ij = e*(p ij - p' ij)

The restricted Boltzmann's strength is it performs a non-linear transformation so it's easy to expand, and can give a hierarchical layer of features. The Weakness is that it has complicated calculations of integer and real-valued neurons. It does not follow the gradient of any function, so the approximation of Contrastive divergence to maximum likelihood is improvised. [14]

Literature

Fischer, Asja; Igel, Christian (2012), "An Introduction to Restricted Boltzmann Machines", Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications , Lecture Notes in Computer Science, vol. 7441, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 14– 36, doi : 10.1007/978-3-642-33275-3_2 , ISBN 978-3-642-33274-6

See also

Autoencoder

Helmholtz machine

References

Bibliography

Chen, Edwin (2011-07-18). "Introduction to Restricted Boltzmann Machines" . Edwin Chen's blog .

Nicholson, Chris; Gibson, Adam. "A Beginner's Tutorial for Restricted Boltzmann Machines" . Deeplearning4j Documentation . Archived from the original on 2017-02-11 . Retrieved 2018-11-15 . {{ cite web }} : CS1 maint: bot: original URL status unknown (link)

Nicholson, Chris; Gibson, Adam. "Understanding RBMs" . Deeplearning4j Documentation . Archived from the original on 2016-09-20 . Retrieved 2014-12-29 .

External links

Python implementation of Bernoulli RBM and tutorial

SimpleRBM is a very small RBM code (24kB) useful for you to learn about how RBMs learn and work.

Julia implementation of Restricted Boltzmann machines: https://github.com/cossio/RestrictedBoltzmannMachines.jl