

Title: Inductive bias

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

PageID: 173926

Categories: Category: Bias, Category: Machine learning

Source: Wikipedia (CC BY-SA 4.0).

The inductive bias (also known as learning bias) of a learning algorithm is the set of assumptions that the learner uses to predict outputs of given inputs that it has not encountered. [1] Inductive bias is anything which makes the algorithm learn one pattern instead of another pattern (e.g., step-functions in decision trees instead of continuous functions in linear regression models). Learning involves searching a space of solutions for a solution that provides a good explanation of the data. However, in many cases, there may be multiple equally appropriate solutions. [2] An inductive bias allows a learning algorithm to prioritize one solution (or interpretation) over another, independently of the observed data. [3]

In machine learning , the aim is to construct algorithms that are able to learn to predict a certain target output. To achieve this, the learning algorithm is presented some training examples that demonstrate the intended relation of input and output values. Then the learner is supposed to approximate the correct output, even for examples that have not been shown during training. Without any additional assumptions, this problem cannot be solved since unseen situations might have an arbitrary output value. The kind of necessary assumptions about the nature of the target function are subsumed in the phrase inductive bias . [1] [4]

A classical example of an inductive bias is Occam's razor , assuming that the simplest consistent hypothesis about the target function is actually the best. Here, consistent means that the hypothesis of the learner yields correct outputs for all of the examples that have been given to the algorithm.

Approaches to a more formal definition of inductive bias are based on mathematical logic . Here, the inductive bias is a logical formula that, together with the training data, logically entails the hypothesis generated by the learner. However, this strict formalism fails in many practical cases in which the inductive bias can only be given as a rough description (e.g., in the case of artificial neural networks), or not at all.

Types

The following is a list of common inductive biases in machine learning algorithms.

Maximum conditional independence : if the hypothesis can be cast in a Bayesian framework, try to maximize conditional independence. This is the bias used in the Naive Bayes classifier .

Minimum cross-validation error : when trying to choose among hypotheses, select the hypothesis with the lowest cross-validation error. Although cross-validation may seem to be free of bias, the "no free lunch" theorems show that cross-validation must be biased, for example assuming that there is no information encoded in the ordering of the data.

Maximum margin : when drawing a boundary between two classes, attempt to maximize the width of the boundary. This is the bias used in support vector machines . The assumption is that distinct classes tend to be separated by wide boundaries.

Minimum description length : when forming a hypothesis, attempt to minimize the length of the description of the hypothesis.

Minimum features : unless there is good evidence that a feature is useful, it should be deleted. This is the assumption behind feature selection algorithms.

Nearest neighbors : assume that most of the cases in a small neighborhood in feature space belong to the same class. Given a case for which the class is unknown, guess that it belongs to the same class as the majority in its immediate neighborhood. This is the bias used in the k-nearest

neighbors algorithm . The assumption is that cases that are near each other tend to belong to the same class.

Shift of bias

Although most learning algorithms have a static bias, some algorithms are designed to shift their bias as they acquire more data. [5] This does not avoid bias, since the bias shifting process itself must have a bias.

See also

Algorithmic bias

Cognitive bias

No free lunch theorem

No free lunch in search and optimization

References

v

t

e

Acquiescence

Ambiguity

Affinity

Anchoring

Attentional

Attribution Actor–observer Fundamental Group Ultimate

Actor–observer

Fundamental

Group

Ultimate

Authority

Automation

Double standard

Availability Mean world

Mean world

Belief

Blind spot

Choice-supportive

Commitment

Confirmation Selective perception

Selective perception

Compassion fade

Congruence

Cultural

Declinism
Distinction
Dunning–Kruger
Egocentric Curse of knowledge
Curse of knowledge
Emotional
Extrinsic incentives
Fading affect
Framing
Frequency
Frog pond effect
Halo effect
Hindsight
Horn effect
Hostile attribution
Impact
Implicit
In-group
Intentionality
Illusion of transparency
Mean world syndrome
Mere-exposure effect
Narrative
Negativity
Normalcy
Omission
Optimism
Out-group homogeneity
Outcome
Overton window
Precision
Present
Pro-innovation
Proximity
Response
Rosy retrospection
Restraint
Self-serving

Social comparison
Social influence bias
Spotlight
Status quo
Substitution
Time-saving
Trait ascription
Turkey illusion
von Restorff effect
Zero-risk
In animals
Estimator
Forecast
Healthy user
Information Psychological
Psychological
Lead time
Length time
Non-response
Observer
Omitted-variable
Participation
Recall
Sampling
Selection
Self-selection
Social desirability
Spectrum
Survivorship
Systematic error
Systemic
Verification
Wet
Academic
Basking in reflected glory
Déformation professionnelle
Funding
FUTON

Inductive
Infrastructure
Inherent
In education
Liking gap
Media False balance Vietnam War South Asia United States Arab–Israeli conflict Ukraine
False balance
Vietnam War
South Asia
United States
Arab–Israeli conflict
Ukraine
Net
Political bias
Publication
System justification
Reporting
White hat
Ideological bias on Wikipedia
Cognitive bias mitigation
Debiasing
Heuristics in judgment and decision-making
Lists: General
Memory
v
t
e
Differentiable programming
Information geometry
Statistical manifold
Automatic differentiation
Neuromorphic computing
Pattern recognition
Ricci calculus
Computational learning theory
Inductive bias
IPU
TPU

VPU

Memristor

SpiNNaker

TensorFlow

PyTorch

Keras

scikit-learn

Theano

JAX

Flux.jl

MindSpore

Portals Computer programming Technology

Computer programming

Technology