



Want help with machine learning? [Take the FREE Crash-Course](#).

Search...

# What is a Hypothesis in Machine Learning?

by [Jason Brownlee](#) on [March 4, 2019](#) in [Start Machine Learning](#)

Supervised machine learning is often described as the problem of approximating a target function that maps inputs to outputs.

This description is characterized as searching through and evaluating candidate hypothesis from hypothesis spaces.

The discussion of hypotheses in machine learning can be confusing for a beginner, especially when “*hypothesis*” has a distinct, but related meaning in statistics (e.g. [statistical hypothesis testing](#)) and more broadly in science (e.g. scientific hypothesis).

In this post, you will discover the difference between a hypothesis in science, in statistics, and in machine learning.

After reading this post, you will know:

- A scientific hypothesis is a provisional explanation for observations that is falsifiable.
- A statistical hypothesis is an explanation about the relationship between data populations that is interpreted probabilistically.
- A machine learning hypothesis is a candidate model that approximates a target function for mapping inputs to outputs.

Let’ s get started.



A Gentle Introduction to Hypotheses in Machine Learning  
Photo by [Bernd Thaller](#), some rights reserved.

## Overview

This tutorial is divided into four parts; they are:

1. What Is a Hypothesis?
2. Hypothesis in Statistics
3. Hypothesis in Machine Learning
4. Review of Hypothesis

## What Is a Hypothesis?

A hypothesis an explanation for something.

It is a provisional idea, an educated guess that requires some evaluation.

A good hypothesis is testable; it can be either true or false.

In science, a hypothesis must be falsifiable, meaning that there exists a test whose outcome could mean that the hypothesis is not true. The hypothesis must also be framed before the outcome of the test is known.

“ ... not any hypothesis will do. There is one fundamental condition that any hypothesis or system of hypotheses must satisfy if it is to be granted the status of a scientific law or theory. If it is to form part of science, an hypothesis must be falsifiable.”

A good hypothesis fits the evidence and can be used to make predictions about new observations or new situations.

The hypothesis that best fits the evidence and can be used to make predictions is called a theory, or is part of a theory.

- **Hypothesis in Science:** Provisional explanation that fits the evidence and can be confirmed or disproved.

## What Is a Hypothesis in Statistics?

Much of statistics is concerned with the relationship between observations.

Statistical hypothesis tests are techniques used to calculate a **critical value** called an *“effect.”* The critical value can then be interpreted in order to determine how likely it is to observe the effect if a relationship does not exist.

If the likelihood is very small, then it suggests that the effect is probably real. If the likelihood is large, then we may have observed a statistical fluctuation, and the effect is probably not real.

For example, we may be interested in evaluating the relationship between the means of two samples, e.g. whether the samples were drawn from the same distribution or not, whether there is a difference between them.

One hypothesis is that there is no difference between the population means, based on the data samples.

This is a hypothesis of no effect and is called the null hypothesis and we can use the statistical hypothesis test to either reject this hypothesis, or fail to reject (retain) it. We don’t say “accept” because the outcome is probabilistic and could still be wrong, just with a very low probability.

“... we develop a hypothesis and establish a criterion that we will use when deciding whether to retain or reject our hypothesis. The primary hypothesis of interest in social science research is the null hypothesis

If the null hypothesis is rejected, then we assume the alternative hypothesis that there exists some difference between the means.

- **Null Hypothesis (H0):** Suggests no effect.
- **Alternate Hypothesis (H1):** Suggests some effect.

Statistical hypothesis tests don’t comment on the size of the effect, only the likelihood of the presence or absence of the effect in the population, based on the observed samples of data.

- **Hypothesis in Statistics:** Probabilistic explanation about the presence of a relationship between observations.

## What Is a Hypothesis in Machine Learning?

Machine learning, specifically supervised learning, can be described as the desire to use available data to learn a function that best maps inputs to outputs.

Technically, this is a problem called function approximation, where we are approximating an unknown target function (that we assume exists) that can best map inputs to outputs on all possible observations from the problem domain.

An example of a model that approximates the target function and performs mappings of inputs to outputs is called a hypothesis in machine learning.

The choice of algorithm (e.g. neural network) and the configuration of the algorithm (e.g. network topology and hyperparameters) define the space of possible hypothesis that the model may represent.

Learning for a machine learning algorithm involves [navigating the chosen space of hypothesis](#) toward the best or a good enough hypothesis that best approximates the target function.

“Learning is a search through the space of possible hypotheses for one that will perform well, even on new examples beyond the training set.

This framing of machine learning is common and helps to understand the choice of algorithm, the problem of learning and generalization, and even the bias-variance trade-off. For example, the training dataset is used to learn a hypothesis and the test dataset is used to evaluate it.

A common notation is used where lowercase-h (*h*) represents a given specific hypothesis and uppercase-h (*H*) represents the hypothesis space that is being searched.

- ***h (hypothesis)*:** A single hypothesis, e.g. an instance or specific candidate model that maps inputs to outputs and can be evaluated and used to make predictions.
- ***H (hypothesis set)*:** A space of possible hypotheses for mapping inputs to outputs that can be searched, often constrained by the choice of the framing of the problem, the choice of model and the choice of model configuration.

The choice of algorithm and algorithm configuration involves choosing a hypothesis space that is believed to contain a hypothesis that is a good or best approximation for the target function. This is very challenging, and it is often more efficient to spot-check a range of different hypothesis spaces.

“ We say that a learning problem is realizable if the hypothesis space contains the true function. Unfortunately, we cannot always tell whether a given learning problem is realizable, because the true function is not known.

— Page 697, [Artificial Intelligence: A Modern Approach](#), Second Edition, 2009.

It is a hard problem and we choose to constrain the hypothesis space both in terms of size and in terms of the complexity of the hypotheses that are evaluated in order to make the search process tractable.

“ There is a tradeoff between the expressiveness of a hypothesis space and the complexity of finding a good hypothesis within that space.

— Page 697, [Artificial Intelligence: A Modern Approach](#), Second Edition, 2009.

- **Hypothesis in Machine Learning:** Candidate model that approximates a target function for mapping examples of inputs to outputs.

## Review of Hypothesis

We can summarize the three definitions again as follows:

- **Hypothesis in Science:** Provisional explanation that fits the evidence and can be confirmed or disproved.
- **Hypothesis in Statistics:** Probabilistic explanation about the presence of a relationship between observations.
- **Hypothesis in Machine Learning:** Candidate model that approximates a target function for mapping examples of inputs to outputs.

We can see that a hypothesis in machine learning draws upon the definition of a hypothesis more broadly in science.

Just like a hypothesis in science is an explanation that covers available evidence, is falsifiable and can be used to make predictions about new situations in the future, a hypothesis in machine learning has similar properties.

A hypothesis in machine learning:

1. **Covers the available evidence:** the training dataset.
2. **Is falsifiable (kind-of):** a test harness is devised beforehand and used to estimate performance and compare it to a baseline model to see if is skillful or not.
3. **Can be used in new situations:** make predictions on new data.

Did this post clear up your questions about what a hypothesis is in machine learning?  
Let me know in the comments below.

## Further Reading

This section provides more resources on the topic if you are looking to go deeper.

### Books

- [What Is This Thing Called Science?](#), Third Edition, 1999.
- [Statistics In Plain English](#), Third Edition, 2010.
- [Artificial Intelligence: A Modern Approach](#), Second Edition, 2009.
- [Machine Learning](#), 1997.

### Posts

- [A Gentle Introduction to Applied Machine Learning as a Search Problem](#)
- [A Gentle Introduction to Statistical Hypothesis Tests](#)
- [Critical Values for Statistical Hypothesis Testing and How to Calculate Them in Python](#)
- [15 Statistical Hypothesis Tests in Python \(Cheat Sheet\)](#)

### Discussions

- [What is hypothesis in machine learning?](#), Quora.
- [What exactly is a hypothesis space in the context of Machine Learning?](#), Cross Validated.
- [What is Hypothesis set in Machine Learning?](#), Cross Validated.

### Articles

- [Hypothesis](#), Wikipedia.
- [Statistical hypothesis testing](#), Wikipedia.

## Summary

In this post, you discovered the difference between a hypothesis in science, in statistics, and in machine learning.

Specifically, you learned:

- A scientific hypothesis is a provisional explanation for observations that is falsifiable.
- A statistical hypothesis is an explanation about the relationship between data populations that is interpreted probabilistically.
- A machine learning hypothesis is a candidate model that approximates a target function for mapping inputs to outputs.