

Search Data Science Ce [Search](#)

- [Sign Up](#)
- [Sign In](#)



# Data Science Central

THE ONLINE RESOURCE FOR BIG DATA PRACTITIONERS

• [HOME](#) [DATAVIZ](#) [HADOOP](#) [BIG DATA](#) [ANALYTICS](#) [WEBINARS](#) [DEEP LEARNING](#) [AI](#) [JOBS](#) [MEMBERSHIP](#) [SEARCH](#) [CLASSIFIEDS](#) [CONTACT](#)

[Subscribe to DSC Newsletter](#)

- [All Blog Posts](#)
- [My Blog](#)
- [Add](#)



## Machine Learning Vs. Statistics

- [Posted by Amelia Matteson on August 18, 2017 at 11:30am](#)
- [View Blog](#)

*This article was written by [Aatash Shah](#).*

## Machine Learning Vs. Statistics



Many people have this doubt, what's the difference between statistics and machine learning? Is there something like machine learning vs. statistics?

From a traditional data analytics standpoint, the answer to the above question is simple.

- Machine Learning is an algorithm that can learn from data without relying on rules-based programming.
- Statistical modeling is a formalization of relationships between variables in the data in the form of mathematical equations.

Machine learning is all about predictions, supervised learning, unsupervised learning, etc.

Statistics is about sample, population, hypothesis, etc.

Two different critters, right? Well, let's see if they are actually that different!

**Both machine learning and statistics have the same objective:**

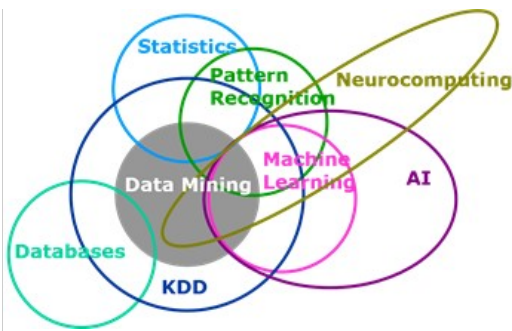
According to Larry Wasserman,

*"They are both concerned with the same question: how do we learn from data?"* In his blog, he states how the same concepts have different names in the two fields,

Statistics		Machine learning
Estimation	~	Learning
Classifier	~	Hypothesis
Data Point	~	Example/ Instance
Regression	~	Supervised Learning
Classification	~	Supervised Learning
Covariate	~	Feature
Response	~	Label

Robert Tibshirani, a statistician and machine learning expert at Stanford, calls machine learning "glorified statistics".

Nowadays, both machine learning and statistics techniques are used in pattern recognition, knowledge discovery and data mining. The two fields are converging more and more even though the below figure may show them as almost exclusive.



Source: [SAS Institute](#); A Venn diagram that shows how machine learning and statistics are related

Both machine learning and statistics share the same goal: learning from data. Both these methods focus on drawing knowledge or insights from the data. But, their methods are affected by their inherent cultural differences.

They're related, sure. But their parents are different.

Machine learning is a subfield of computer science and artificial intelligence. It deals with building systems that can learn from data, instead of explicitly programmed instructions.

A statistical model, on the other hand, is a subfield of mathematics.

Machine learning is comparatively a new field.

Cheap computing power and availability of large amounts of data allowed data scientists to train computers to learn by analyzing data. But, statistical modeling existed long before computers were invented.

#### Methodological differences between machine learning and statistics:

The difference between the two is that machine learning emphasizes optimization and performance over inference which is what statistics is concerned about.

This is how a statistician and machine learning practitioner will describe the outcome of the same model:

ML professional: "The model is 85% accurate in predicting Y, given a, b and c."

Statistician: "The model is 85% accurate in predicting Y, given a, b and c; and I am 90% certain that you will obtain the same result."

Machine learning requires no prior assumptions about the underlying relationships between the variables. You just have to throw in all the data you have, and the algorithm processes the data and discovers patterns, using which you can make predictions on the new data set. Machine learning treats an algorithm like a black box, as long it works. It is generally applied to high dimensional data sets, the more data you have, the more accurate your prediction is.

In contrast, statisticians must understand how the data was collected, statistical properties of the estimator (p-value, unbiased estimators), the underlying distribution of the population they are studying and the kinds of properties you would expect if you did the experiment many times. You need to know precisely what you are doing and come up with parameters that will provide the predictive power. Statistical modeling techniques are usually applied to low dimensional data sets.

To read the full original article click [here](#). For more machine learning statistics related articles on DSC click [here](#).

#### DSC Resources

- Services: [Hire a Data Scientist](#) | [Search DSC](#) | [Classifieds](#) | [Find a Job](#)
- Contributors: [Post a Blog](#) | [Ask a Question](#)
- Follow us: [@DataScienceCtrl](#) | [@AnalyticBridge](#)

#### Popular Articles

- [Difference between Machine Learning, Data Science, AI, Deep Learnin...](#)
- [What is Data Science? 24 Fundamental Articles Answering This Question](#)
- [Hitchhiker's Guide to Data Science, Machine Learning, R, Python](#)
- [Advanced Machine Learning with Basic Excel](#)

Views: 14199

Like

7 members like this

Share

Tweet

 Facebook

 Like 0

- [< Previous Post](#)
- [Next Post >](#)

## Comment

You need to be a member of Data Science Central to add comments!

[Join Data Science Central](#)



Comment by [David Reinke](#) on August 24, 2017 at 7:17pm

This is somewhat oversimplified in my view. The comparison refers to *frequentist* statistics, not Bayesian statistics. Vapnik's book *The Nature of Statistical Learning Theory* seems to me to present a more accurate comparison.

[RSS](#)

Welcome to  
Data Science Central

[Sign Up](#)  
or [Sign In](#)

Or sign in with:

- 
- 
- 
- 
- 

**FOLLOW US**

[@DataScienceCtrl](#) | [RSS Feeds](#)

## TOP CONTENT



1 [Machine Learning – Can We Please Just Agree What This Means](#)



2 [Understanding Neural Network: A beginner's guide](#)



3 [How to measure customer satisfaction correctly](#)



4 [Trump Solves Famous Mathematical Conjecture](#)



5 [What is Data Science? 24 Fundamental Articles Answering This Question](#)



6 [Twelve types of Artificial Intelligence \(AI\) problems](#)

- [RSS](#)
- [View All](#)

## ANNOUNCEMENTS

[DSC Complimentary Classifieds Listings Extended \(through Dec 31, 2017\)](#)

[\[eBook\] A Unified Approach to Analytics with Apache Spark](#)

[Agile Data Mastering - Complimentary O'Reilly Report](#)

[Advanced, open source data analytics: Webinar](#)

[Advance Your Career: NYU Stern MS in Business Analytics](#)

[Advanced Analytics for Data Scientists](#)

[Marketing Attribution with Automated Machine Learning](#)

[\[Last chance\] Best Price for Strata 2018 ends Friday](#)

[Your Springboard to a Data Science Career](#)

[Cloud BI strategies from leading analyst firm](#)

## VIDEOS



• [Advanced Tool for Dynamic Visual Analytics](#)

Added by [Tim Matteson](#) 0 Comments 0 Likes



• [Deep Learning - From Basic Principles through Training Models for Deployment into Production](#)

Added by [Tim Matteson](#) 1 Comment 6 Likes



•

[DSC Webinar Series: Powerful, Flexible and Accessible Code-free Data Science](#)

Added by [Andrei Macsin](#) 0 Comments 0 Likes

- [Add Videos](#)
- [View All](#)

## RESOURCES

- [Migrating an Excel Spreadsheet to MySQL and to Spark 2.0.1 \(Part 1\)](#)
- [Introduction to Programming in Stata](#)
- [Benchmarking 20 Machine Learning Models Accuracy and Speed](#)
- [Stata Cheat Sheet](#)
- [Selection of best articles from our past weekly digests](#)
- [Statistical Analysis Advisor Chart](#)
- [Selection of best articles from our past weekly digests](#)
- [Free Online Book: Forecasting, Principles and Practice](#)
- [38 Seminal Articles Every Data Scientist Should Read](#)
- [Black-box Confidence Intervals: Excel and Perl Implementation](#)

## TOP CATEGORIES

[Machine Learning](#)

[R Programming](#)

[Python for Data Science](#)

[Visualization, Dashboards](#)

[NoSQL and NewSQL](#)

[Big Data](#)

[Cheat Sheets](#)

[Internet of Things](#)

[Excel](#)

© 2017 Data Science Central Powered by **NING**

[Badges](#) | [Report an Issue](#) | [Privacy Policy](#) | [Terms of Service](#)