AI and Social Science – Brendan O'Connor
cognition, language, social systems;
statistics, visualization, computation

Statistics vs. Machine Learning, fight!

Posted on December 3, 2008

10/1/09 update — well, it's been nearly a year, and I should say not everything in this rant is totally true, and I certainly believe much less of it now. Current take: *Statistics*, not machine learning, is the real deal, but unfortunately suffers from bad marketing. On the other hand, to the extent that bad marketing includes misguided undergraduate curriculums, there's plenty of room to improve for everyone.

So it's pretty clear by now that statistics and machine learning aren't very different fields. I was recently pointed to a very amusing comparison by the excellent statistician — and machine learning expert — Robert Tibshiriani. Reproduced here:

Glossary	
Machine learning	**Statistics**
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = $1,000,000	large grant = $50,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

Hah. Or rather, ouch! I had two thoughts reading this. (1) Poor statisticians. Machine learners invent annoying new terms, sound cooler, and have all the fun. (2) What's wrong with statistics? They have way less funding and influence than it seems they might deserve.

There are several issues going on here, both substantive and cultural:

There might be too much re-making-up of terms on the ML side. But lots of these are useful. "Weights" is a great, intuitive term for the parameters of a linear model. I use it all the time to explain classifiers and regressions to non-experts. I was surprised to see "test set" on the statistics side; I'm used to thinking of held-out test set accuracy as an extremely common ML technique, while in statistics model fit is assessed with parametric assumptions for standard errors and such. I really like cross-validation and bootstrapping as ways of thinking about generalization — again, something that's far easier to grasp than sampling and hypothesis testing approaches to parameter inference — which keep getting taught to and misunderstood by generations of confused Introduction to Statistics students. For example, how many times has been explained that: No, a p-value is NOT the probability your model is wrong. But scientific papers regularly treat significance levels in that manner (look how many stars are on this result!) On the other hand, cross-validation accuracy *is* something you can interpret as being related to the probability your model is right.

I'll also note that there are definitely a number of topics in ML that aren't very related to statistics or probability. Max-margin methods: if all we care about is prediction, why bother using a probability model at all? Why not just optimize the spatial geometry instead? SVM's don't require a lick of probability theory to understand. (Of course probability-based approaches are huge in ML, but it's important to remember they're not the only game in town, and there is no necessary reason they must be.) And then there are non-traditional settings such as online learning, reinforcement learning, and active learning, where the structure of access to information is in play. There are certainly plenty of things in statistics that aren't considered part of ML — say, regression diagnostics and significance testing. Finally, many ML problems involve large, high dimensional data and models, where computational issues are very important. For example, in statistical machine translation, alignment models are described with probability theory and fit to data, but their structure is complex enough that optimal inference is intractable, and how you do approximate inference (EM, Viterbi, beam search, etc.) is a very major issue.

But the most interesting differences between stats and ML are institutional.

I've been hearing lots of friends compare two dueling courses at Stanford: CS229, the CS department's "machine learning" course taught by Andrew Ng; and Stat 315 A/B, the Statistics department's "statistical learning" sequence taught by some combination of Tibshirani, Jerome Friedman, and Trevor Hastie. These people are all top-of-the-line researchers in the field. Their courses' contents are extremely similar; I'd bet any of them could teach most of the material from the other side.

What differs most is the teaching style. CS has far better lecture notes. Of course, the stats people wrote a very good book; but better lecture notes win because I can access them later and send them to people for free. CS students I've talked to think the CS course is better taught; I can't find stats students who take the CS course. (My sample is biased, though I know people in both.) Finally, the CS course has a big, open-ended project component; the Stats course follows more of a traditional problem set and tests format.

I think this is reflective of the differences in institutional culture between CS and Stats. There's an interesting John Langford post on part of the issue, which he calls "The Stats Handicap". He points out that stats Ph.D.'s have a big disadvantage in the job market because statistics has an old-school journal-oriented publishing culture, so students publish much less and have less experience engaging with a research community. CS is conference-oriented — certain conferences have a higher prestige than many journals (e.g. NIPS in ML, CHI in HCI) — and this results in faster turnaround, dissemination, and collaboration. (I've heard others make similar comparisons between CS and psychology.) I'd expect any discipline with a larger conference emphasis to have better courses since they should reward presentation/teaching skills — or at least encourage practice — more than in journal world.

ML sounds like it's young, vibrant, interesting to learn, and growing; Stats does not.

Is marketing a problem? Machine learning terms definitely sound pretty cool. Maybe the perspective of computational intelligence lends itself to cool names. Though the Stanford statisticians certainly know how to play this game — for example, they made up their own names for variants of L1 and L2-regularized regression, leaving annoyed people like me forever googling "lasso" and "ridge" trying to remember which is which. (On the other hand, perhaps that's child's play compared to the true original sin of ML nomenclature: tossing around the highly deceptive term "neural network" for a stack of linear functions paired with a wonky, overhyped training algorithm; the combination of which, many years later, still causes confusion. Definitely blame CS for that one.)

Another issue is the definition of statistics itself. In 1997, Jerome Friedman wrote an extremely interesting analysis of the situation: "Data Mining and Statistics: What's the Connection?". He points out, quite correctly, the statistical impoverishment of some common approaches to data mining. You can certainly blame statistics for not marketing its ideas well enough, or blame CS for ignoring statistics. For example there's a good case that lots of genetic algorithms and neural network research was much ado about nothing — that is, over-complicated cool-sounding hammers looking for nails when all you needed were some time-honored statistical and optimization

techniques. (E.g. why NN when you haven't tried a straight-up GLM? Why GA when you haven't tried Nelder-Mead?) But this problem has been rectified somewhat — for example, NLP has seen a big move to simple linear models as the default technique, and NN's and GA's have fallen from grace in mainstream ML.

Friedman argues part of the problem is in how statisticians approach problems and the world:

> One can catalog a long history of Statistics (as a field) ignoring useful methodology developed in other data related fields. Here are some of them that had seminal beginnings in Statistics but for the most part were subsequently ignored in our field: Pattern Recognition, Neural Networks, Machine Learning, Graphical Models, Chemometrics, Data Visualization.

That is not to say statistics is not important — it's incredibly important. He quotes Efron as saying "Statistics has been the most successful information science." However, information science is becoming bigger and broader and more exciting, thanks to computation and ever-increasing amounts of data. What should statisticians do? Friedman continues (light editing and emphasis is mine):

> One view says that our field should concentrate on that small part of information science that we do best, namely probabilistic inference based on mathematics. If this view is adopted, we should become resigned to the fact that the role of Statistics as a player in the "information revolution" will steadily diminish over time.
>
> Another point of view holds that statistics ought to be concerned with **data analysis**. The field should be defined in terms of a set of problems — rather than a set of tools — that pertain to data. Should this point of view ever become the dominant one, a big change would be required in our practice and academic programs.
>
> First and foremost, we would have to make peace with computing. It's here to stay; that's where the data is. This has been one of the most glaring omissions in the set of tools that have so far defined Statistics. Had we incorporated computing methodology from its inception as a fundamental statistical tool (as opposed to simply a convenient way to apply our existing tools) many of the other data related fields would not have needed to exist. They would have been part of our field.

Friedman wrote this article more than 10 years ago. All his observations about the importance and increasing prevalence of data and computing power are even more true today than back then. Has the field of statistics changed? Not clear. (I'd appreciate seeing evidence to the contrary.)

On the other hand a world of data *has* to be increasingly statistical. The positive spin from Efron:

> A new generation of scientific devices, typified by microarrays, produce data on a gargantuan scale – with millions of data points and thousands of parameters to consider at the same time. These experiments are "deeply statistical". Common sense, and even good scientific intuition, won't do the job by themselves. Careful statistical reasoning is the only way to see through the haze of randomness to the structure underneath. Massive data collection, in astronomy, psychology, biology, medicine, and commerce, is a fact of 21st Century science, and a good reason to buy statistics futures if they are ever offered on the NASDAQ.

I know that I'm interested in quantitative information science, including statistics and data analysis. Machine learning has many strengths, but it is definitely an odd way to go about analysis. But there's a good case that statistics, as traditionally defined, is only going to have a smaller role in the future. "Data mining" sounds more relevant, but does it even exist as a coherent subject? Maybe it's time to study a more applied statistical field like econometrics.

This entry was posted in Best Posts. Bookmark the permalink.

132 Responses to *Statistics vs. Machine Learning, fight!*

Pingback: *Machine learning — The Endeavour*

Pingback: *Brendan O'Connor puts machine learning and statistics in a jar and shakes the jar « Mike Love's blog*

Carlos *says:*
December 3, 2008 at 10:19 pm

Somewhat related to this:
http://tinyurl.com/breiman2001
Leo Breiman, Statistical Modeling: The Two Cultures.
From the abstract
"The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets."

brendano *says:*
December 4, 2008 at 1:58 am

Very interesting, very related paper; thanks.

ekzept *says:*
December 4, 2008 at 4:23 am

To paraphrase Feynman, something (some problem, some theorem, some assessment, some analysis, some method) isn't new just because it has been given a new name.

miked98 *says:*
December 4, 2008 at 6:34 am

Cool post. It seems to be a general phenomenon what happens when applied sciences (psychology, biology, economics) adopt methods of the [relatively more] pure sciences (physics, statistics, mathematics), or when newer disciplines (synthetic biology) venture into space occupied by an older one (electrical engineering).

In the long run, things do get cleaned up. But it may take a younger generation of thinkers, with a stronger grasp of the pure sciences (think of the mathematicians who've ventured into economics, or the physicists who've gotten into synthetic biology) to sort out the mess of nomenclature and integrate the fields.

Pingback: *Machinel Learning İstatistiğe Karşı: Dövüş Başlasın! | FZ Blogs*

Bob Carpenter *says:*
December 4, 2008 at 6:08 pm

Tibshirani's graph should really have included two additional factors: (1) average number of courses taught/year, and (2) median student/post-doc/faculty stipends and salaries. I think it's part of the explanation of grant size, since most CS folks I knew at CMU simply bought themselves out of teaching by having grants pay their salaries. This makes CS departments much more highly leveraged (grant money needed per person to sustain the operation; at CMU our budget from the university didn't even cover tenured faculty salaries, much less T.A.s). Even so, a $1M grant in CS is going to support a lot more people than a $50K grant in stats. You see an even stronger form of this effect in medical research, which has huge salaries and huge grants with armies of cheap post-docs.

Once you put average courses taught/year and see the statisticians with 3 or 4 and the machine learning people with 1 or maybe 2

Maybe we just need larger grants so we can go to the Alps and pay expensive graduate students and get away without having to actually teach for a living. I think the big tell would be

Bob Carpenter *says:*

December 4, 2008 at 6:38 pm

Blame the physicists for the term "max entropy", which is just plain old logistic regression (as are one-layer neural networks with sigmoid activations or softmax). But the non-Bayesian statisticians get the blame for "regularization", and "lasso"/"ridge"; they're just priors to the Bayesians.

Don't diss back-prop! It's having a renaissance in stochastic gradient methods all over machine learning.

Folks in machine learning are discovering Bayesian methods of dealing with uncertainty, whereas the Bayesian statisticians been using graphical models in custom and general-purpose systems like BUGS for decades.

Dan Jurafsky and I were just discussing ML vs. stats, because we've both been doing more social science type stats. We were both surprised, like Brandon, that the statisticians don't use cross-validation. I speculated it's largely because the statistical paradigm of evaluating fit leads them to build models focused on analyzing existing data sets rather than to doing forward-looking predictions, but there are lots of counterexamples, such as FiveThirtyEight, which predicted the 2008 U.S. presidential election very accurately using Bayesian methods over polls.

The other issue Dan and I discussed is that statisticians care deeply about their coefficients (weights, parameters, whatever), whereas machine learning folks tend to toss them all into a bin and let priors and cross-validation sort them out. Sure, we might look at the feature weights to make sure the algorithm's doing something sensible, but we don't write papers where the point is to explore the effect of the word "the" (a feature) on estimation (there actually should be more of these papers in ML, in my opinion). For instance, statisticians very much want to explore the effect of a person's weight on their chance of diabetes and aren't going to be very happy giving a doctor an SVM and saying "trust it, it worked well on cross-validation". And they want to examine the role of income or church attendance on voting. The goal is to explore the parameters ("effects") as much as predict which way a state's going to vote in the next election.

Finally, let me point out that the main systems used for microarrays in practice are simple linear factor models that any statistician would recognize, like dChip. What's the justification for Efron's comment that "Careful statistical reasoning is the only way to see through the haze of randomness to the structure underneath."? Does statistics imply probability? If not, what about SVMs, as Brandon asks? Maybe all we need is room-sized 3D visualization coupled to human brain power.

brendano *says:*

December 4, 2008 at 10:03 pm

On descriptive statistics by attention to coefficients — lots of social science empirical work involves small, limited situations where they're trying to find out if certain effects are in play; extrapolation to other situations is usually done with reasoning by analogy. If your reasoning and decision making in future situations is going to be qualitative, a trained-up SVM from a different situation isn't useful; but knowing the top 3 coefficients from a linear model there *is* useful qualitative information.

I think this is the point of that bit about the jeff hammerbacher talk we were discussing at http://anyall.org/blog/2008/07/the-macgyver-of-data-analysis/ — he's assuming the domain of analyzing web behavior logs and figuring out how to make a website better. you could worry about automated decision making for what content to show people (ranking, recommendations etc.); but probably the most productive thing to do is extract qualitative insights from the data to inform the design process. this is a pretty social science-y domain; t-tests and linear regressions are going to be the tools of choice.

brendano *says:*

December 7, 2008 at 9:39 am

Another response, from Andrew Gelman — on a rather pro-CS note:
http://www.stat.columbia.edu/~cook/movabletype/archives/2008/12/machine-learnin.html

I wish a statistician would come here aggressively defend their discipline. At the very least — what about experimental design? Or tricky low-evidence situations: don't you want a statistician, not an MLer, to testify at a trial about the whether an event was a coincidence?

Ethan Bauley *says:*
December 8, 2008 at 12:50 am

Have you ever studied game theory?

---

brendano *says:*
December 8, 2008 at 5:42 am

yes, why?

Pingback: *Statistics vs. Machine Learning vs. Data Mining, fight! » No Random Walking!*

Will Dwinnell *says:*
December 11, 2008 at 3:33 am

"…the highly deceptive term "neural network" for a stack of linear functions paired with a wonky, overhyped training algorithm; …"

The term "neural network" covers a broad range of techniques, but I don't think the above description accurately describes any of them. For one thing, any "stack of linear functions" reduces algebraically to a single linear function. I imagine that you are referring to a multi-layer perceptron, but that is built of a stack of non-linear functions.

-Will Dwinnell
Data Mining in MATLAB

---

brendano *says:*
December 11, 2008 at 3:44 am

I did mean a multilayer perceptron. Individual units are "linear" in the generalized linear models sense — the response is a function of a linear combination of the inputs. (The same way a logistic regression is a linear model; you stack a them up to get a multilayer NN.) Trained with backpropagation this is theoretically very powerful, but unfortunately is tricky to use in practice. Thus "overhyped."

Sorry for any confusion. And nice blog, by the way.

Ethan Bauley *says:*
December 11, 2008 at 5:34 pm

@brendano

I was just asking because of your comment about econometrics. I just discovered game theory a couple of months ago and have been reading some books about applications to business strategy, like Nalebuff/Brandeberger's "Co-opetition." Great stuff; seems applicable to the things you're interested in.

I met a CalTech PhD on a flight back from San Jose a couple of weeks ago who just finished his degree in theoretical computer science (emphasis on game theory). He interviewed at YHOO and they were looking for algorithms that identify Nash equilibria in huge data sets of user interactions.

fwiw

;-)

---

lilly *says:*

December 16, 2008 at 1:32 am

On the conference vs journal oriented cultures, one frustration I have with conference oriented cultures is that they still feel too slow and competitive for getting feedback on work but so fast that they encourage the publication of a lot of low hanging fruit work so you can be up on a podium every year (or multiple times a year, depending on how many prestigious conferences are in your field).

How do people decide what is a valid contribution for presentation at NIPS? I can't seem to make sense of it at CHI, except for a particularly traditional sort of "build system – evaluate on 10 research lab mates – summarize results" that ends up not creating very powerful or synthetic new knowledge.

brendano *says:*

December 30, 2008 at 9:51 pm

Hey Lilly, I don't really know what the NIPS criteria are. They do both theory and applied papers. I do know that John Langford has a bunch of interesting things to say about it and in general about review criteria.

http://hunch.net/?p=499
http://hunch.net/?p=191
http://hunch.net/?p=223

brendano *says:*

December 30, 2008 at 9:52 pm

Ah, he has a number of interesting posts here: http://hunch.net/?cat=33

Tyrone Carson *says:*

January 8, 2009 at 11:51 pm

hi
afgbiiiq50qhstbe
good luck

Pingback: *Comparison of data analysis packages: R, Matlab, SciPy, Excel, SAS, SPSS, Stata - Brendan O'Connor's Blog*

Pingback: *Statistics vs Machine Learning « Data Mining, a Course by Blog*

David Warde-Farley *says:*

March 4, 2009 at 8:52 am

So, the "stack of linear functions" I took issue with too. A neural network with any modeling power has the crucial ingredient of nonlinearities at each hidden layer. The way I think about a single layer neural network is as a logistic regression model operating on a set of features where the feature extraction is learned as well. As for "overhyped" = tricky to use, yes, that's true. There are entire books written on effectively training the damned things. Unlike SVMs, you need to know a little something about both your domain problem and the model in order to get good results. There's been a recent renaissance, as Bob points out, in neural networks research with the advent of methods for training deep networks (which is nearly impossible with gradient descent + backprop alone, unless you tie a lot of the weights together i.e. convolutional networks).

The basic dichotomy between statistics and machine learning that I see is in academic lineage. Yes, they've invented new names for lots of things, but that's mostly because the machine learning community grew out of computer scientists, engineers, physicists (I am often taken aback at just how many physicists seem to pop up), and yes, theoretical neuroscientists back in the 1980s, with very little crosstalk with statistics. There's a strong difference in problem focus, as you mention.

I also think your discussion centers somewhat unfairly on classification and regression. There's plenty of interesting work being done in unsupervised learning of complex, generative models of data, both with prior knowledge built in and without. Pleasantly, the two communities have converged on graphical models as a common parlance for describing probabilistic models of data.

Abhijit *says:*

March 5, 2009 at 3:23 am

Just discovered this post. Fantastic!! It aligns with many of my thoughts, specially since I'm a biostatistician interested in high-dimensional problems where ML techniques seem to be "easier". Still learning about ML methods, though. You're right about the cross-validation bit, though. Statisticians aren't necessarily trained in predictive modeling and their techniques, including CV, model averaging, bagging, … I've recently felt the need for learning these areas since they're apropos of some problems I'm consulting on. There NEEDS to be more cross-fertilization of the two fields, since we keep re-inventing wheels.

Hanif *says:*

March 13, 2009 at 8:46 am

A great blog and interesting discussion – I'm an engineer who's spent the last decade in biotech and pharma. Whenever someone non-analytical asks what I do, I say "data mining" which is not far from the truth. However, more recently I've spent more time with statistical modeling and the associated community in biostatistics.

I definitely see the differences Brendan and Bob mention – the focus on understanding the factor parameters/effects has a lot to with the fact that the same analysts helped design the study which includes specifying which data to collect and contrasts to select. Many of the machine learning folks seem to be more contract mercernaries / collaborating scientists who came on post-study.

The Netflix contest seems to be an interesting context to compare the approaches and insights gained. The highest performing [most?] groups seem to be dominated by ML. My guess is that Netflix will favor the white-box modelers to hel them decide how to modify their actual suggestion engine, as opposed to wrap around the best performer's algorithm.

My experience is in biological data, which is almost always underspecified and overdetermined, full of correlated variables (the correlation itself being informative), and answers used as a starting point of another study/experiment/analysis. A ML-informed statistical (or statistically-informed ML) modeling approach ends up being the most useful. I use R mostly, with a lot of other domain-specific tools, and am starting to use MATLAB more as well (deployment of visualizations).

Keep up the great articles…

brendano *says:*

May 1, 2009 at 3:09 am

Abhijit –

> Statisticians aren't necessarily trained in predictive modeling
> and their techniques, including CV, model averaging, bagging…

I agree and have thought the same thing myself, but it's still funny to read that sentence given that there are so many papers from stats journals on those topics. I think they were all, or most of them, invented by statisticians too. Maybe this is showing a difference between statistics proper versus applied stats in traditional biology and social science.

But sometimes the ideas are there just named differently. Economists know about the held-out accuracy method of evaluation (e.g. CV & friends). They call it "out of sample predictions". Economists, of course, broke away hard from mainline stats a while ago, calling it "econometrics" and reinventing names for EVERYTHING, plus throwing in a bizarre obsession with the method of moments. In terms of intellectual arrogance and needless renaming/duplication, economists are much worse than computer scientists and engineers. Maybe as bad as physicists.

Ping Li *says:*

August 30, 2009 at 1:53 pm

Very nice post and interesting discussions. However, your statement of "I can't find stats students who take the CS course" is indeed biased.

I am a junior faculty in Statistics, with a Ph.D. in Statistics (advised by Prof. Hastie). I have master's degrees in CS, EE (both from Stanford), and other fields. I interned as software engineer (RealNetworks Server Team and Microsoft Visual Studio) and later spent many summers at MSR.

I totally do not like to use small datasets as I (and possibly many others) believe results on small datasets could often be fairly easily tuned and one can hardly test the signficance on results from small datasets. However, I sometimes find I must also use small datasets, since they were used in many (CS) machine learning papers.

I am just one example of "statistician". There are much better examples. The statistics folks at ATT and Google are doing wonderful adorable things.

There are probably some non-statisticians who are used to view "statisticans" as "statisticans". My research proposals often received comments like "this is just a statistican's view of ..." " this work is not inter-disciplinary" etc.

I really believe there should not be any "statisticians' view" or "Computer Scientiststs' view" As long as the algorithms work on the real data, then it is a good view. Why should we bother putting a label?

Pingback: *My daily readings 09/29/2009 « Strange Kite*

jimmy *says:*
October 15, 2009 at 3:41 am

http://www-stat.stanford.edu/~hastie/Papers/ESLII.pdf

Brendan O'Connor *says:*
October 15, 2009 at 3:43 am

Great book. I posted that to a class mailing list earlier this week then today suddenly it's all over programming websites like reddit. I always wonder...

Pingback: *"Statistics vs. Machine Learning, fight!" « Trying to Make Sense of Data*

Bastian *says:*
December 27, 2009 at 7:19 pm

Hi guys,
you seem to be experts. I am a student in economics and I am working with Markov Logic Networks.
But now I have a question: in machine learning, there are no parameters but weigths. In econometrics an important thing is to look, if there is a significant influence, how strong it is and in which direction it goes. From weigths (at least in Markov Logic Networks) I can't get this information. So how can this problem be solved? By estimating the probability distribution and compare the result with a random probability distribution (how it is done in graph theory)?
For examining empirical results and for testing theories I'm not sure, if machine learning is superior against statistics.
What is your opinion.

With regards.

brendano *says:*
December 27, 2009 at 7:26 pm

Hi Bastian,

I, at least, am not an expert at all!

Weights are exactly the same thing as parameters. In fact, weights in an MLN are very similar to parameters in a logit regression. (Similarly with other log-linear structured models like CRFs and MRFs.) But their interpretation might be a little more complex given the structuredness of MLNs.

MLNs have their own mailing list that might be useful to try your question on; "alchemy-discuss", I think it's called. Somewhere on the UWashington website.

You're right that techniques developed in ML-land are usually less focused on making accurate descriptive inferences. I've never seen anyone try to do significance testing for MLNs, for example. I think this is a big weakness of ML, at least as it's usually conceived. All this stuff will be merged together eventually, but in the meantime, there's still confusion.

Bastian *says:*
December 29, 2009 at 7:56 pm

Hello Brendan,

thank you for the answer! Wrigth, I know "alchemy-discuss", and I posed there some theoretical questions. But nobody answered, I don't know why.

Bye bye.

Rhiannon weaver *says:*
January 12, 2010 at 11:01 pm

At Carnegie Mellon stats, we've been aware of this for quite some time. I started there in 2000 and one of our first semester courses was stat computing. with the prevalence of bayesian methods (whereby you CAN figure out the 'probability your model is right'), and practical ways of estimating complex hierarchical models, you have to take a very problem-oriented approach. See this letter to the American Statistician by Kass and Brown:

http://pubs.amstat.org/doi/abs/10.1198/tast.2009.0019

I would also argue, however, the ML doesn't say very much at all about experimental design and/or controlling for multiple sources of error in experiments. You mention above a lot of things that ML has that stats doesn't; there's one thing at least that stats has that ML doesn't.

---

pm *says:*
April 8, 2010 at 11:56 pm

Hi Brandon,

I stumbled across your blog about a year ago and peeked back in every now and then since. Considering myself a probabilist who comes from the theoretical side, I would not really call myself a statistician, but I am certainly more open to statistical methods than to machine learning.

I got in touch with machine learners after leaving university, and the one thing that puzzled me most and is most critical to me is at the very heart of what modeling means. To me, ML is more focused on methods and techniques, and less on concepts that are suited to the problem at hand. The culture of ML is METHOD oriented, not PROBLEM ORIENTED, as it seems to me.

To me, a model is anything that describes the important parts of a real world phenomenon I am interested in. Networks or graphs are only certain instances, or examples, of what a model might constitute. Model choice is super-critical in any data analysis you carry out, and any statistical inference and/or prediction you carry out is only valid within the model you chose. Moreover, by using a 'tool', you choose a model implicitly, always. There are no exceptions to this rule.

A model might be given by a graph, a stochastic differential equation, a specification of distributional assumptions etc. The set of statistical methods that are suitable when observing data which are supposed to be generated by the model dynamics follows from the model assumptions. Many statistical methods are standard nowadays, and they are often employed without really asking whether the underlying assumptions are true. For example, even when using standard software and doing such trivial things as calculating sample means, you make assumptions about your data. (In this case, you assume the data are sufficiently independent and identically distributed.)

And this is what people from the ML community seem not to be aware of. By choosing a neural network, an SVM, or any other kind of super-flexible mechanism and fitting that to your data, they make the assumption that the data are generated by the dynamics the tool implies.

The model is implied by the tool, the tool _replaces_ the model.
Depending on the application, the consequences of this approach
are more or less serious. And sometimes they are very serious…
e.g. in financial engineering when heavy tailed phenomena are of
paramount importance, but ML techniques are mostly based on the
assumption that noise follows a Gaussian law…

In my opinion, this is what really makes up the different cultures
between ML and statistics. Good statisticians are well aware of the
limitations of their tools, MLers aren't… what do you think?

brendano *says:*
April 9, 2010 at 1:07 am

@pm, that sounds about right to me. I think the best ML theory and practices are turning into more of the statistical-style approach, of
understanding both the power and limits of the techniques in question.

Ping Li *says:*
May 26, 2010 at 1:01 pm

@pm.

Regarding your comments on "ML techniques are mostly based on the
assumption that noise follows a Gaussian law…"

Successful (for example, in industry) ML methods such as trees (together with boosting) are not affected much by the heavy-tailed nature of the
data.

You mentioned SVM, which has rich and beautiful theories. My limited experience is that SVM works extremely well (and very fast) when the
data are "nice" (such as MNIST). As soon as the data become "difficult", the performance of SVM may drop dramatically (can more experienced
folks correct me on this?). We academia researchers love SVM-type of algorithms because we have the time and passion to carefully tune the
parameters, designing kernels (if one kernel does not work, use multiples), clean the data (remove "outliers"), normalized the data, etc.

(I should add that linear SVM seems to be the right tool when the data are extremely high-dimensional, sparse, more or less binary, for
example, text data).

My guess is that both ML and statistics folks are well aware of the limitations, but one might be often under the pressure (for example, publish
or perish) of developing sophisticated algorithms that may work well only on a few (and often small or even contrived) datasets but may not
work well in general. The current publication model in ML seems to favors sophisticated stuff. Just my humble opinion.

Akshay Bhat *says:*
September 1, 2010 at 12:46 am

A post written in 2008 bashing ANN's is really pointless, why don't you
talk about Probabilistic Graphical Models, Support Vector Machines or other area of learning such as Unsupervised or Online or
Reinforcement Learning. Plus you don't discuss Vapnik's Statistical Theory of learning or PAC theory? Nor do you mention Deep learning
architectures like restricted Boltzmann machines. What about emerging problem in Networks and Link Prediction. Or even whole subfield of
Recommendation Systems [Collaborative Filtering as some call it?].

If all those terms sound too much, then what tools [not invented by CS researchers] does statistics currently posses to deal with problems such
as Reinforcement Learning?

If you are going to bash Machine Learning by using ANN (Popular 1995 – 2005) and back-propagation (popular circa 1990′s) in 2008, isn't
convincing.

Discounting Machine Learning by calling it as Statistics is saying all Biology is Chemistry and all Chemistry is Physics.

cttet *says:*
December 22, 2012 at 6:04 am

Graphical Models and Support Vector Machines in my view are quite stats..

But I wonder why everyone's definition of the terms are different.

For people I know, those are called statistical learning.

But NN, GA and other methods are different.

Akshay Bhat *says:*
September 1, 2010 at 12:49 am

I take back my comment I didn't read the text properly,

Pingback: *Machine learning hay Statistics « MFEPE*

Pingback: *Learning about Machine Learning | Honglang Wang's Blog*

Pingback: *Statistics vs. Machine Learning, fight! | Honglang Wang's Blog*

Pingback: *Statistics vs. Machine Learning, fight! | Honglang Wang's Blog*

Pingback: *Comparison of data analysis packages: R, Matlab, SciPy, Excel, SAS, SPSS, Stata | Honglang Wang's Blog*

Pingback: *Quora*

Pingback: *Applying social psychology | Ready-to-hand*

Arthur *says:*
August 25, 2011 at 4:00 pm

Hello!

I've just landed on this blog post and found it very interesting.

We're nearly two years since your last update. Has your position evolved? Statistics VS ML, what's the score?

Thanks,

Arthur

Brendan O'Connor *says:*
August 25, 2011 at 6:55 pm

I'm more into stats these days. But I think the gap between the disciplines keeps narrowing anyways.

Pingback: *Quora*

Me Me Me *says:*
November 10, 2011 at 8:01 am

Is there any textbook or so that you would recommend to CS students who have been exposed to ML techniques, but not classical stats ? I would like to read more of that, but I wouldn't know where to start, and if it is approachable for the average CS student ;-)

Brendan O'Connor *says:*
November 10, 2011 at 1:37 pm

@"Me Me Me": Get "All of Statistics" by Larry Wasserman. It's basically written exactly for this use case.

Pumbaa *says:*

November 15, 2011 at 12:32 am

@Brendan Thx for that time your explained Gibbs sampling to me. This post is really interesting!

@ Me Me Me: I highly recommend "All of Statistics", too. If by any chance you are from CMU, I would recommend Larry's "10705 : Intermediate Statistics", too. I am a CS background like you and didn't take much stat courses, but after taking that I feel more equipped for a lot of ML stuff.

Pingback: *[How do I become a data scientist? | spider's space](#)*

Pingback: *[How do I become a data scientist? « Victor Fang's Computing Space](#)*

Pingback: *[Statistical Modeling versus Machine Learning « Data Meaning...](#)*

Pingback: *[Bombarded with big data , big science and big learning « Big AI Dream](#)*

William Payne *says:*

December 17, 2012 at 11:13 am

Well, here at the coal-face, I never really saw much distinction between the two. Machine Learning? Statistic + Algorithms, as far as I am concerned. I pull in what I need at the time that I need it, irrespective of where it originates from in the academic sphere. Having said that, statisticians (and mathematicians too, for that matter) need to pull their finger out when it comes to communicating. Neither I, nor my colleagues have the time nor spare mental capacity for navel-gazing "look-how-clever-I-am" papers. Save the proofs for the appendices. We need clearly presented advice, written in a well-developed tutorial style, wherever possible pitched at an intelligent-early-postgraduate level.

plancherel *says:*

May 20, 2013 at 2:33 am

haha... sorry to come across this comment so long after it was posted.

As a mathematician in a research lab, I cannot tell you how many times I come across people like you. You try an algorithm, that you don't understand, on data, that you don't understand, and you get a bad result, that you don't understand. Then, you ask someone like me or a statistician, to explain why your approach doesn't work.

I usually refer you guys to some paper, that you won't understand, and then put my finger back in it. :-)

cttet *says:*

December 22, 2012 at 6:01 am

Statistial learning is of course related to statistics.

But you cannot desregard NN and GA in two sentence. Machine learning is far more general!

Pingback: *[Data Science | ModrnWiki (Pre-Alpha)](#)*

Anders Nielsen *says:*

March 7, 2013 at 12:22 am

I have a good foundation on applied stats and mathematical statistics.

Could you suggest a book about machine learning for people who knows statistics?

Thanks!

Kind regards,

Anders

Brendan O'Connor *says:*

March 7, 2013 at 12:26 am

I like this book generally, the new Murphy textbook. http://www.cs.ubc.ca/~murphyk/MLbook/index.html

Also great is Hastie et al, and it is free! http://www-stat.stanford.edu/~tibs/ElemStatLearn/

Pingback: *How do I become a data scientist? | i4igeeks*

Ron Kenett *says:*

May 12, 2013 at 4:16 am

My solution to this is to encourage the development of a Theory of Applied Statistics
See: http://ssrn.com/abstract=2171179

Al DeLosSantos *says:*

May 24, 2013 at 7:45 pm

Thanks for a great post Brendan, very helpful. I took Andrew Ng's ML course on Coursera last fall and can highly recommend him and the course if anyone wants to learn some fundamental ML methods. Great lectures and well prepared assignments that introduce the ML methods using the Octave environment. All throughout the course I kept asking myself (should have posted this in the discussion group!) how I could reconcile his material with what I had previously studied in my few Statistics courses. Your blog discussion has helped…I just have to keep learning and participating in the discussion. :^)

Pingback: *What is Machine Learning | Machine Learning Mastery*

---

voyance sans attendre *says:*

June 3, 2014 at 3:44 pm

consultation astrologie en ligne parisDecouvrir ce excellent site web : voyance par telephone gratuite

---

Pingback: *Unrelated to all that, 6/26 | neuroecology*

Pingback: *機械学習とは何か？ – 機械学習の定義と、使える言い回し | POSTD*

Pingback: *유전자 프로그래밍과 트레이딩 첫째*

Pingback: *How do I become a data scientist? | lordtomriddle*

Pingback: *Le web et ses mécanismes « naturels » | CaddE-Réputation*

Pingback: *How can I become a data scientist? | Data Scientist Wanna Be*

Pingback: *[repost]How do I become a data scientist? | Intelligence Computing*

Pingback: *如何成为数据科学家？来自quora：How can I become a data scientist? | 数据化学*

Pingback: *Statistics vs Machine Learning: The two worlds | The Data Scientist*

Pingback: *All I do is complain, complain … | Hypergeometric*

Pingback: *How do i become a Data Scientist – Primose Training*

Pingback: *Categories of Machine Learning Algorithms – Learning the Machines*

Pingback: *The fight between STAT and ML – TAIJI☺*

Pingback: *Sự khác nhau giữa Học máy, Thô´ng kê và Khai phá dữ liệu | MFEPE*

Pingback: *数据分析师的基本素养——论如何成为一名数据科学家 – 王际桥的博客*

Pingback: *How to become a data scientist Part 1 - Soso Blog knowledge share*

Pingback: *10 Steps to Becoming a Data Scientist – DL Recruiting*

Pingback: *pinboard October 9, 2017 — arghh.net*

Pingback: *The Close Relationship Between Applied Statistics and Machine Learning | NEURALSCULPT.COM*

Pingback: *The Close Relationship Between Applied Statistics and Machine Learning – AiProBlog.Com*

Pingback: *The Close Relationship Between Applied Statistics and Machine Learning – Signal Surgeon*

Pingback: *The Close Relationship Between Applied Statistics and Machine Learning - TrendXnow*

Pingback: *統計、機械学習、データマイニング | プログラミングQA.com*

Pingback: *응용 통계와 머신러닝간의 밀접한 관계 - 네피리티*

Pingback: *Statistics vs Machine Learning: The two worlds - TDS*

Pingback: *Statistics vs System Finding out: The 2 worlds -*

Pingback: *Statistics vs Machine Studying: The 2 worlds - Tech 4K*

Pingback: *Statistics vs Machine Learning: The two worlds - Cash AI*

Pingback: *Statistics vs Machine Studying: The 2 worlds - batatachop*

Pingback: *Statistics vs Machine Studying: The 2 worlds - ainewslatest.com*

Pingback: *Statistics vs Machine Studying: The 2 worlds - My Blog*

Pingback: *Statistics vs Machine Studying: The 2 worlds | Channel969*

Pingback: *Statistics vs Machine Studying: The 2 worlds - 2023cricketworldcup*

Pingback: *Statistics vs Machine Learning: The two worlds – StorageNewsBox*

Pingback: *Statistics vs Machine Learning: The two worlds - TC Technology News*

Pingback: *Statistics vs Machine Studying: The 2 worlds - CreatorTechs. All rights reserved.*

Pingback: *Statistics vs Machine Studying: The 2 worlds - ser-artificial.com*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Abdul Ali Tech](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds | Infinity Fact](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Iphones Ringtone Free Download](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Web and Traffic](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - PinSystem](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - TechnologyBuzz](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - techwithtrends.com](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - superpcparts.com](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - digitalwebgeek](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - RRIVER](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - NEXT LEVEL INNOVATION](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Weekly Tidings](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - technologywalk.com](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Tech Talk Central Trenzy Tech](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - hightechnews](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - techsparks4u](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - artificialintelligence360](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - attkley1403](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds -](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Techknology](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Techmaggie](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - WIKHOST](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds - Porfect Life](#)*

Pingback: *[Statistics vs Machine Learning: The two worlds - MyMentorsWorld](#)*

Pingback: *[Statistics vs Machine Studying: The 2 worlds](#)*

AI and Social Science – Brendan O'Connor

8/27/24, 10:47 AM

Statistics vs. Machine Learning, fight! | AI and Social Science – Brendan O'Connor

17/17