Difference between Machine Learning & Statistical Modeling performance

Definition

Machine Learning is …

an algorithm that can learn from data without relying on rules-based programming.

Statistical Modeling is …

Formalization of relationships between variables in the form of mathematical equations.

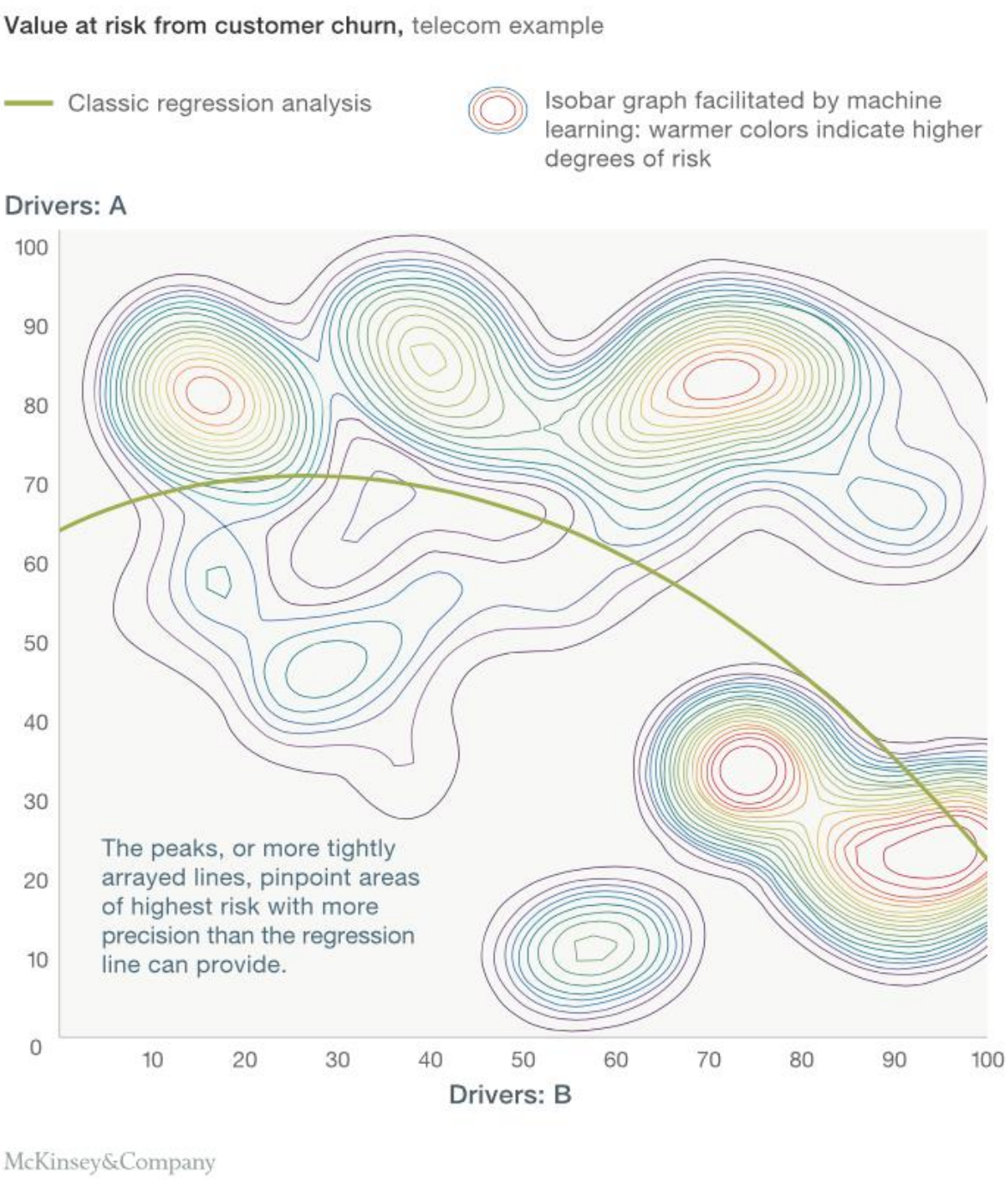
A Business Case

Let us now see an interesting example published by McKinsey differentiating the two algorithms :

Case : Understand the risk level of customers churn over a period of time for a Telecom company

Data Available : Two Drivers – A & B

What McKinsey shows next is an absolute delight! Just stare at the below graph to understand the difference between a statistical model and a Machine Learning algorithm.



. What did you observe from the above graph? Statistical model is all about getting a simple formulation of a frontier in a classification model problem.

Here we see a non linear boundary which to some extent separates risky people from non-risky people. But when we see the contours generated by Machine Learning algorithm, we witness that statistical modeling is no way comparable for the problem in hand to the Machine Learning algorithm. The contours of machine learning seems to capture all patterns beyond any boundaries of linearity or even continuity of the boundaries. This is what Machine Learning can do for you

If this is not an inspiration enough, machine learning algorithm is used in recommendation engines of YouTube / Google etc. which can churn trillions of observations in a second to come up with almost a perfect recommendation. Even with a laptop of 16 GB RAM I daily work on datasets of millions of rows with thousands of parameter and build an entire model in not more than 30 minutes. A statistical model on another hand needs a supercomputer to run a million observation with thousand parameters.