Partial Dependence Plots

In Towards Data Science. More on Medium.

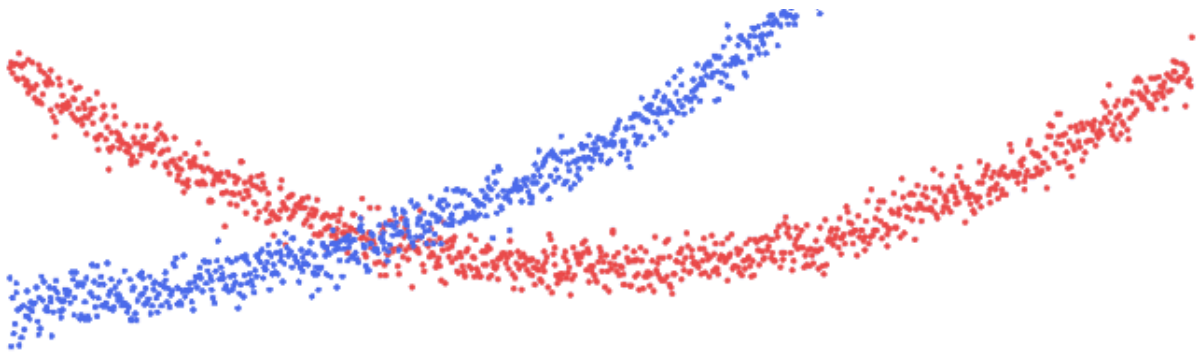
Conor O'Sullivan · Jun 16 ★

HANDS-ON TUTORIALS

Finding and Visualising Non-Linear Relationships

Analysing non-linear relationships with partial dependence plots (PDPs), mutual information and feature importance

When you first start driving you are less experienced and, sometimes, more reckless. As you age, you gain more experience (and sense) and it becomes **less** likely that you're involved in an accident. However, this trend won't continue forever. When you reach old age your eyesight may deteriorate or your reactions may slow. Now, as you age, it becomes **more** likely that you're involved in an accident. This means the probability of an accident has a non-linear relationship with age. Finding and incorporating relationships like these can improve the accuracy and interpretation of your models.



Source: Author

Satya Pattnaik · Oct 22, 2020

Pitfalls To Avoid while Interpreting Machine Learning-PDP/ICE case

Yirui Feng · Oct 7, 2020 ★

How to peek inside a black box model — Understand Partial Dependence Plot

Use partial dependence plot to reveal the effect of targeted features to a black box model.

Photo by Leio McLaren (@leiomclaren) on Unsplash

Abstract

One can find numerous articles today on Explainable AI, some of which can be found **here.** The most standard guide for Explainable AI will undoubtedly be **this book** by **Christoph Molnar.** When I came across the recent paper **Pitfalls to Avoid when Interpreting Machine Learning Models,** I decided to write some blogs out of it. This is

Photo taken by author

Sarit Maitra · Jul 2, 2020 ★

In this post, we will be learning a tool to reveal the working mechanism of a black box model. But before we start, let talk about something else.

EXPLAINING BLACK BOX ALGORITHMS

Explaining ML Black-Box Model

Partial Dependence Plot and SHapley Additive exPlanations

The idea behind is very similar to how we wash food before eating, just like what shows below:

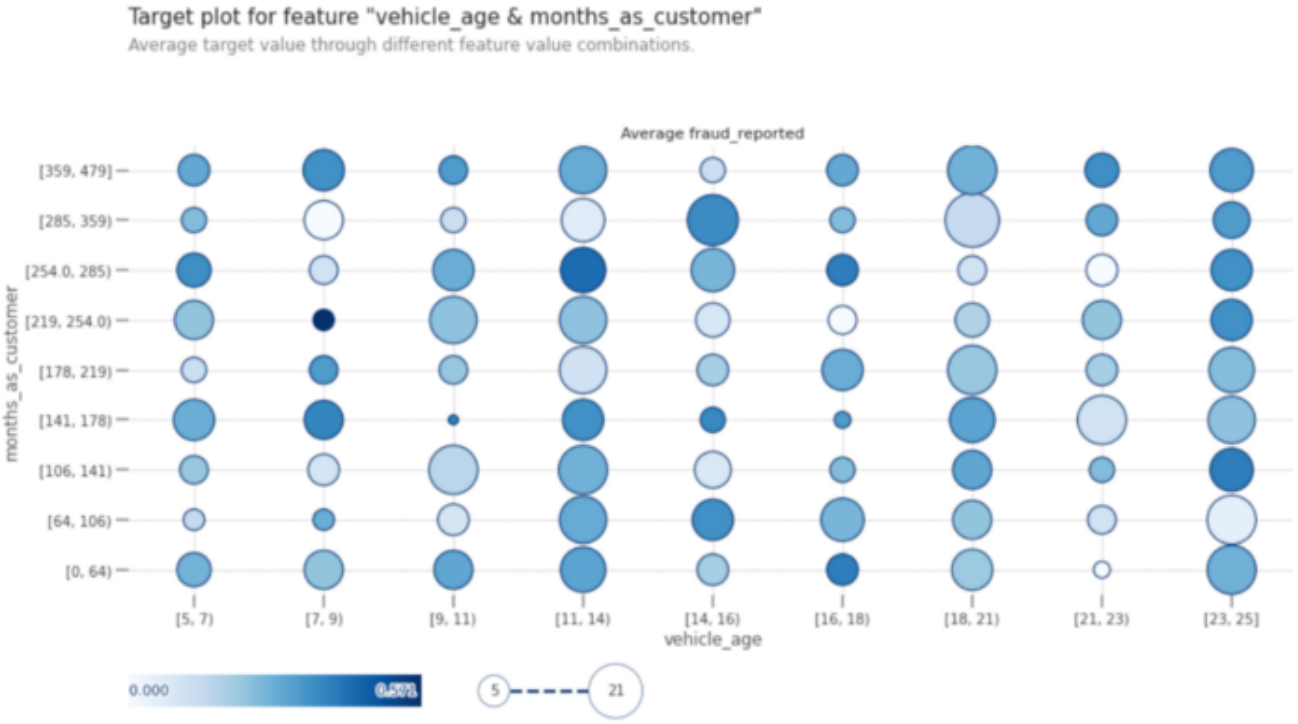


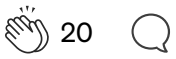
Image by Author

The biggest criticisms of modern ML is about it's black-box nature— input data, output as in prediction/forecasting and that is ML algorithm's workings in simple sense. Here comes the complexity — when we have a highly non-linear model with high degrees of interactions, how can we possibly hope to have a simple understanding of what the

in high risk applications e.g. healthcare as well as high data sensitive areas e.g banking & financials.

Let us revisit…

Read more · 11 min read



Manu Joseph · Nov 16, 2019

Interpretability: Cracking open the black box — Part II

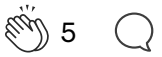
Mean Decrease in Impurity, permutation importance? LOOC importance, PDP plots and more

Is the paywall bothering you? Click here to go around it.

In the last post in the series, we defined what interpretability is and looked at a few interpretable models and the quirks and 'gotchas' in it. Now let's dig deeper into the post-hoc interpretation techniques which is useful when you model itself is not transparent. This resonates with most real world use cases, because whether we like it or not, we get better performance with a black box model.

Income is a pretty popular dataset…

Read more · 21 min read



Werlindo Mangrobang · Jun 6, 2019 ★

What's in the Black Box?

A distillation of "Causal Interpretations of Black-Box Models" by Zhao and Hastie.

Dan Becker · Mar 8, 2019

Why Model Explainability is The Next Data Science Superpower

Photo by Christian Fregnan on Unsplash

Model explainability techniques show you what your model is learning, and seeing inside your model is even more useful than most people expect.

I've interviewed many data scientists in the last 10 years, and model explainability techniques are my favorite topic to distinguish the very best data scientists from the average.

Some people think machine learning models are black boxes, useful for making predictions but otherwise unintelligible; but the best data scientists know techniques to extract real-world insights from any model. For any given model, these data scientists can easily answer questions like

- What features in the data did the model think are most important?