Machine Learning Explainability

Some applications in medical settings

Pei Liu

April 19, 2019

Department of Computer Science @UESTC

Table of contents

Introduction

What is it

Models you built can explain how the event of interest is affected.

- VIP (What variables most affect it?)
- PDP (How do the variables affect it?)
- CASE STUDIES (For an individual?)

If I were in *a linear world*, then how does it become?

$\hat{Y} = Prob(The\ fruit\ is\ an\ apple) = 0.4 \cdot Color + \cdots + 0.2 \cdot Shape + 0.1 \cdot Size$

Why should we do it

But actually, we are building a complex model via Machine Learning methods.

In the realm of supervised learning, there are

- Decision Tree
- Neural Network
- And others

Especially in the medical settings, it is crucial to understand how the factors affect the survival of patients.

Let us focus on the prognostic predictive model that is derived from **West China Hospital**!

Just Explore it!

Details of the model

More details in the IDFS model:

- X variables: Age, Education, ..., Tumor Size. (No. 19)
- Y varibales: Follow-up time and status of relapse. (No. 2)
- Predictions: Hazard Ratio (a real number indicating risk of relapse).
- Algorithm: Cox Proportional Hazard model based on XGBoost.

Since it is indeed a `Black Box`, how can we interpret it just like a linear model?

Variable Importance

What varibles most affect the relapse?

Recommended methods are as follows:

- Built-in functions
- Permutation Importance
- Other tricks?

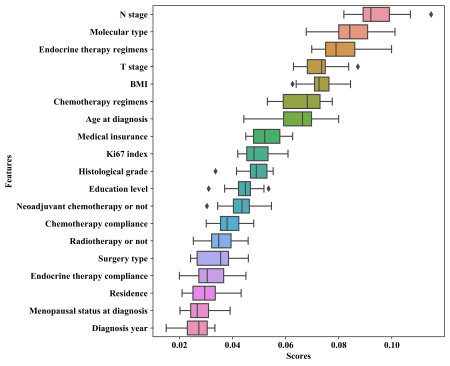

Figure 1: Variable Importance of the IDFS Model

How do the variables affect the relapse?

Partial Dependency Plots

The method we discuss here is **Partial Dependency Plots**.

> *This method takes a row of the dataset and repeatedly changes a value for one feature.*
> *It is done multiple times with different rows and then aggregated in order to find out how the feature is influencing the target on a wide range.*

In order to investigate the effect of a feature, now we exploit it to our IDFS model!

Partial Dependency Plots for A Feature

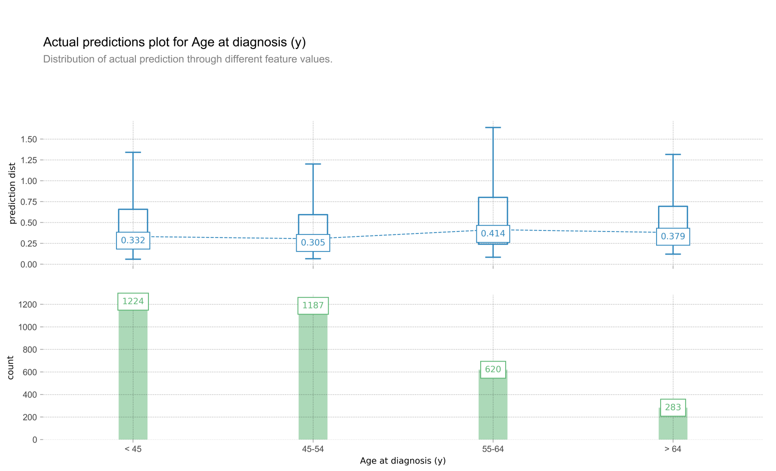

Figure 2: The Actual Prediction Distribution

Partial Dependency Plots for A Feature

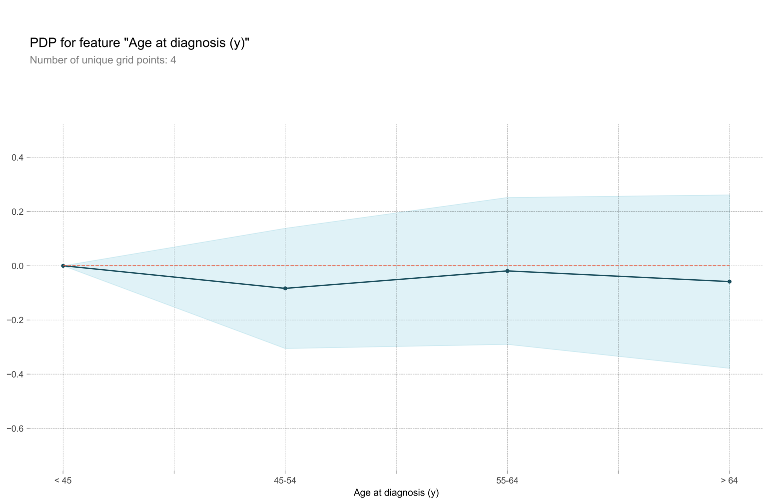

Figure 3: PDP.

Partial Dependency Plots for A Feature

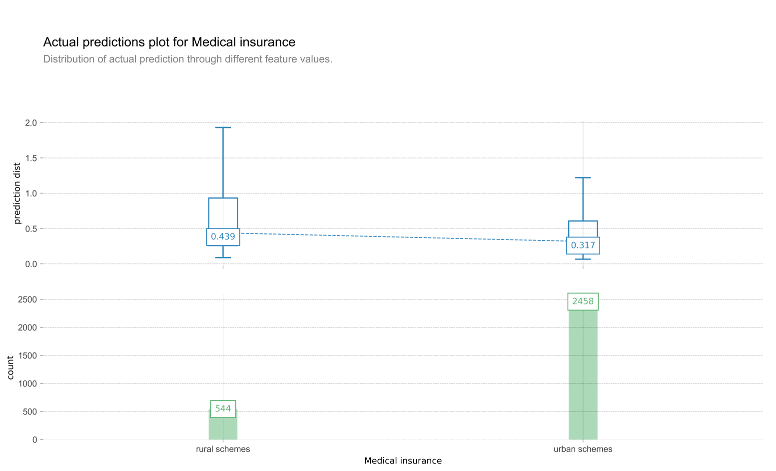

Figure 4: The Actual Prediction Distribution

Partial Dependency Plots for A Feature

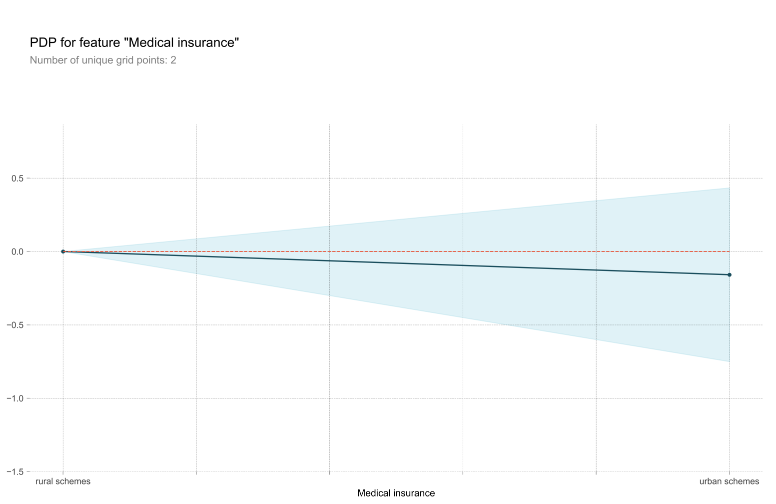

Figure 5: PDP.

Partial Dependency Plots for A Feature

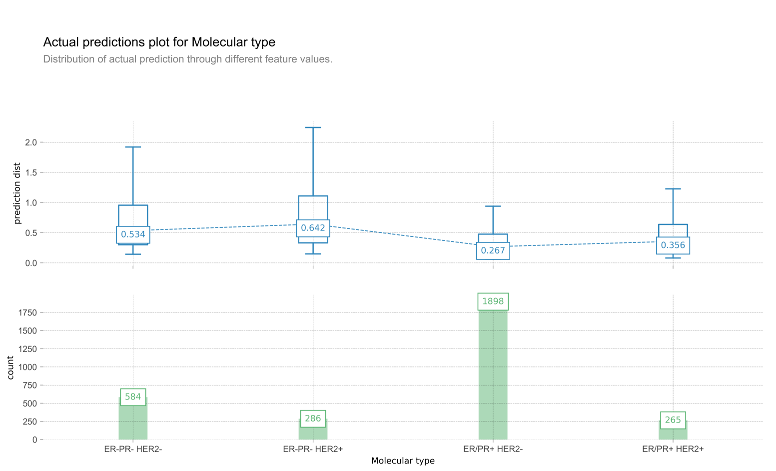

Figure 6: The Actual Prediction Distribution

Partial Dependency Plots for A Feature

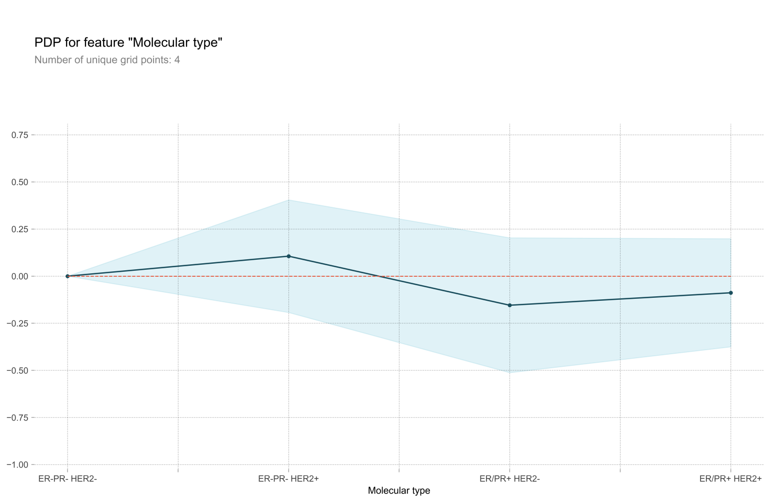

Figure 7: PDP.

Partial Dependency Plots for A Feature

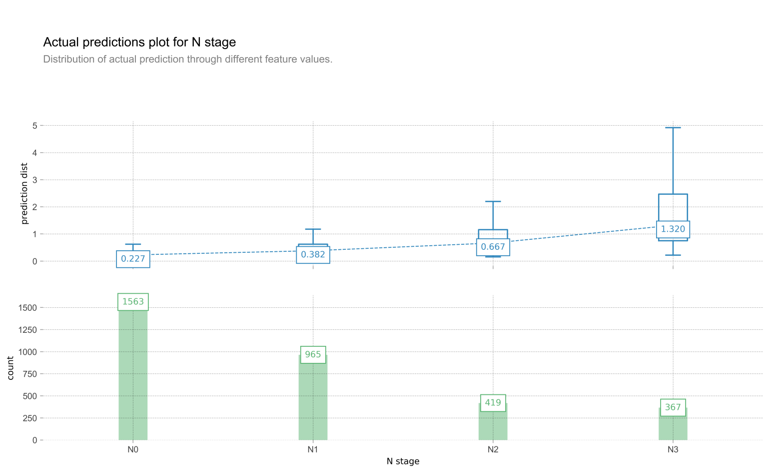

Figure 8: The Actual Prediction Distribution

Partial Dependency Plots for A Feature

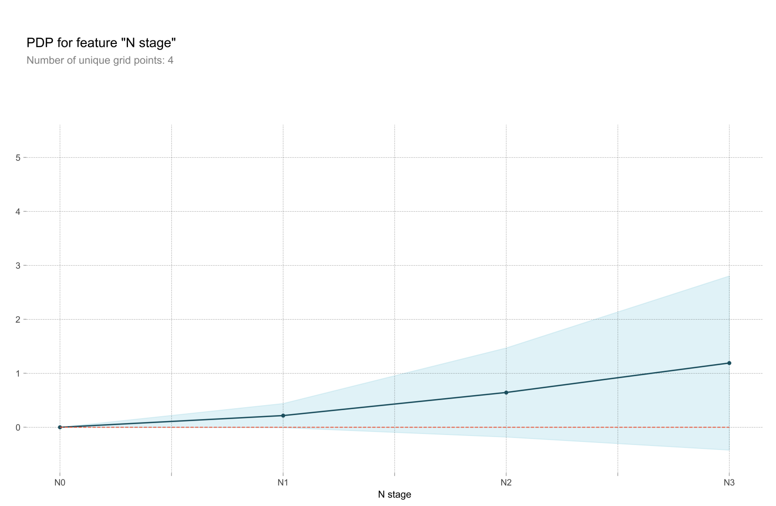

Figure 9: PDP.

Partial Dependency Plots for A Feature

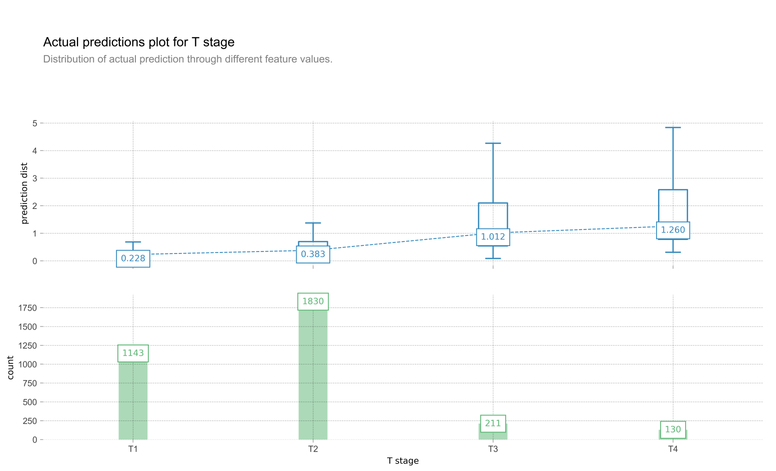

Figure 10: The Actual Prediction Distribution

Partial Dependency Plots for A Feature

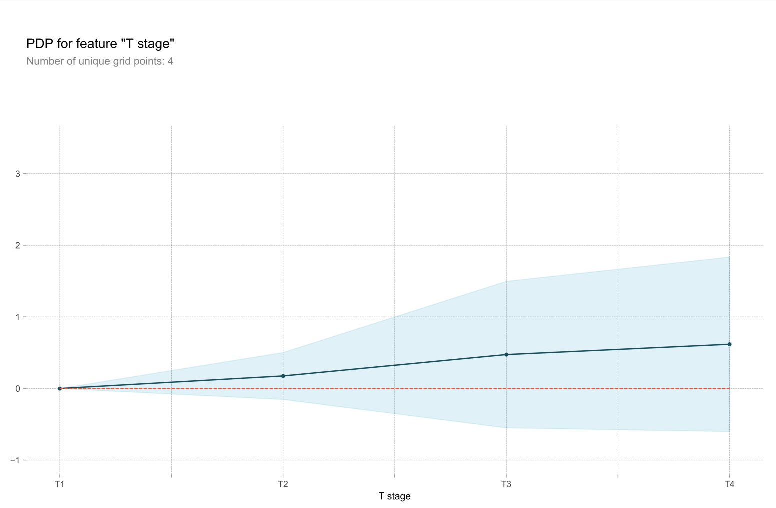

Figure 11: PDP.

Partial Dependency Plots for A Feature

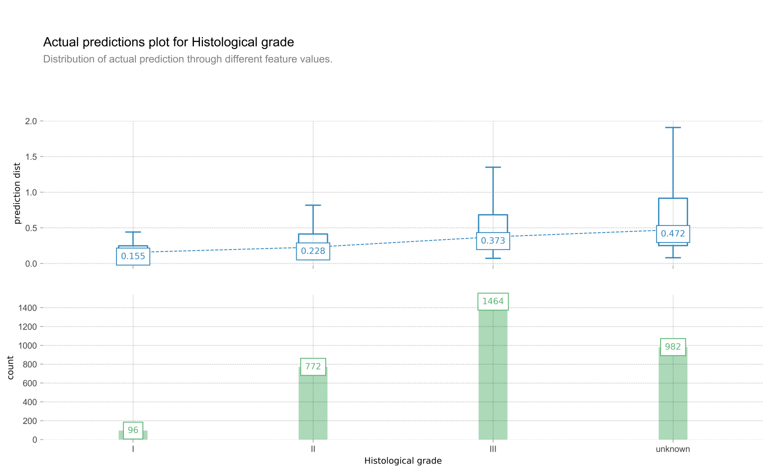

Figure 12: The Actual Prediction Distribution

Partial Dependency Plots for A Feature

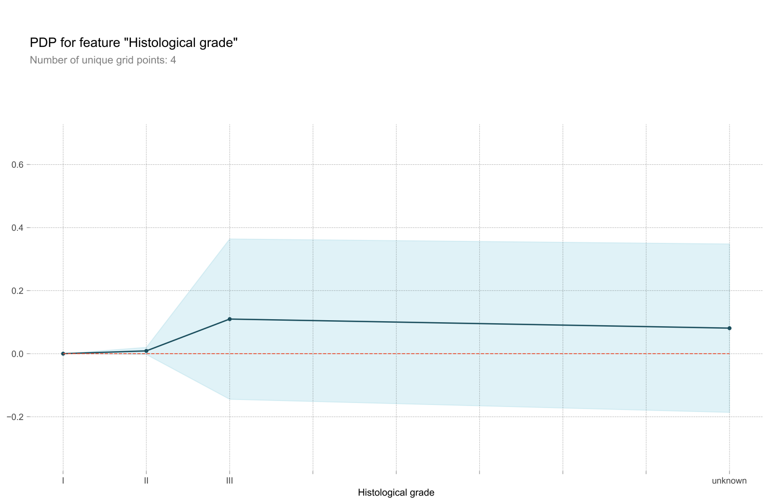

Figure 13: PDP.

Disadvantages of Partial Dependency Plots

- **`Maximum number of features`**: up to 2 features.
- **`Feature distribution`**: you might overinterpret regions with almost no data.
- **`Assumption of independence`**: feature correlations (ridiculous combination of height and weights).
- **`Heterogeneous effects might be hidden`**: PD plots only show the average marginal effects.

Case Studies

How did the variables affect the relapse of a specific patient?

The SHAP values are used to show the effects of the features of a single patient.

Here, the method also takes one feature and compares the value to a baseline value for that feature without changing the other features. That is done for all features.

Need more thinking for the principle!

How did the variables affect the relapse of a specific patient?

An example for showing parameters of a specific patient influencing the risk of hazard ratio.

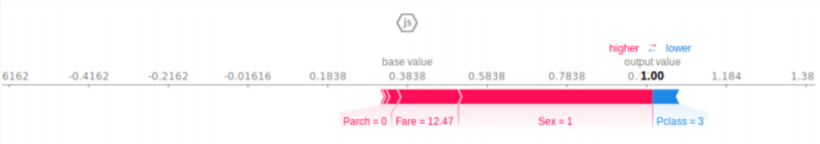

Figure 14: Case Studies

Questions?

References

References:

- Good Books: https://christophm.github.io/interpretable-ml-book/agnostic.html
- Inspired by the tutorial: https://pan.baidu.com/s/1hF6tVG3JvzML0OR4ozZkYQ
- Permutation importance: https://doi.org/10.1093/bioinformatics/btq134
- PDPBOX library: https://github.com/SauceCat/PDPbox
- BEAMER template for presentation: https://github.com/matze/mtheme