Partial Dependency Plots

Jeff Allard
Data Scientist
5/3 Bank
November 2017

1^{st} Motivation: Accuracy is not enough

- In Industry, a predictive model being shown to have strong accuracy* is necessary but seldom sufficient
  - In a regulated industry, there needs to be Compliance / Legal review to safeguard against discrimination / bias
  - Business owner will often want to gain some understanding of how the model makes predictions
  - Sanity check that the model is following accepted relationships between x->y
- General(ized) Linear Models are easy to interpret but seldom demonstrate the strongest accuracy
  - Even these models can become uninterpretable with splines, transformations, interactions etc
- Machine learning models often demonstrate the strongest accuracy but can be very difficult (impossible) to (fully) interpret

*generally, whatever metric is being optimized

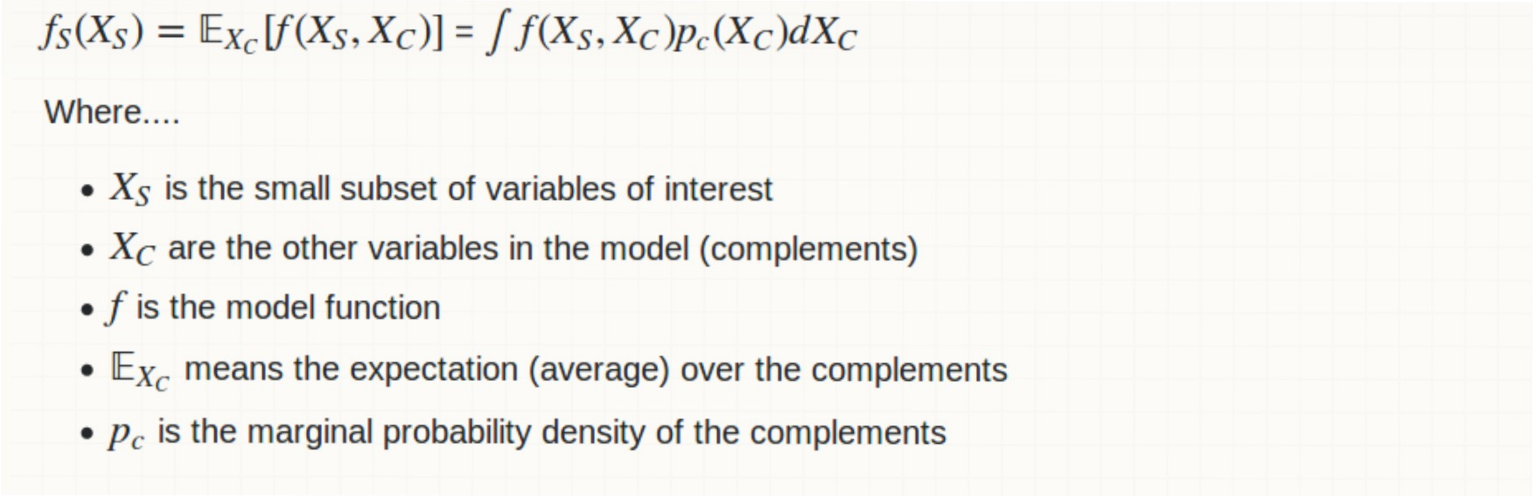
2nd Motivation: Variable Importance is not enough

- Variable importance – a ranking of the most important variables to the model
  - Present in most tree based methods
  - Can be estimated for any model by repeatedly fitting and dropping of variable(s), measuring the impact on performance
- VI is definitely useful but can be misleading
  - No generally accepted definition, even in tree based models
    - # of times was a splitter
    - Gain in information value from split
    - etc
- VI is not fully sufficient
  - Often does not explain how a variable impacts the prediction, just that it does...in some way….often in a very complex way

Partial Dependency Plots (PDP) -Defined-

- We want to understand how the model prediction changes as a predictor(s) within the model changes, given the other variables present

  Formally,



$$f_S(X_S) = \mathbb{E}_{X_c}[f(X_S, X_C)] = \int f(X_S, X_C)p_c(X_C)dX_C$$

Where....

- X_S is the small subset of variables of interest
- X_C are the other variables in the model (complements)
- f is the model function
- $\mathbb{E}_{X_C}$ means the expectation (average) over the complements
- p_c is the marginal probability density of the complements

Each value of the PD function is the expected value of the model estimate when X_s = s (i.e. the value of X_s is fixed) where the expectation is taken over the values of X_c (X_c allowed to vary over its marginal distribution).

Partial Dependency Plots (PDP) - Applied-

In practice, the algorithm to **estimate** this function is very simple

For all values of interest of a given variable(s) – normally some quantiles of the variable (min, 5th percentile, ….median…..95th percentile, max)

1. Create a copy of the training data
2. Set the value of the variable(s) of interest (X_s) to a constant value
3. Predict the outcome using the model over all the training rows (X_c varies)
4. Calculate the mean of the outcome
5. Set aside the pair: {value of X_s, average outcome}

Plot these pairs

Works with ANY supervised learning regression or classification model:

Logistic regression to boosted trees to deep neural networks

DEMO

* Only works for binary classification and regression problems with a predict function

*Only works for numeric variables

When can PDP fail?

We are averaging over all the other variables in the model….

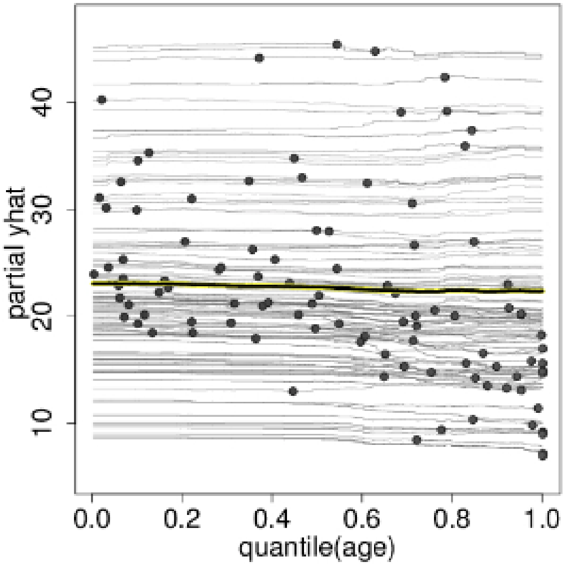
- If there are strong interactions or dependencies between the variable of interest and others in the model
- If we extrapolate outside of the range of values of the variable of interest
- If we create impossible combinations of data not seen in the training data

Methods to Understand Important Interactions

- Domain knowledge
- Specific modeling tools
  - Small scale bivariate testing with logistic / linear regression
  - Interaction scores from GBDT fits (e.g xgboost, catboost)
  - Glinternet : lasso based linear model for interaction variable selection (high dimension)

(Some) Alternatives

- Ice Plots
  - Individual plots of the various complement variable values, instead of average of them (https://arxiv.org/pdf/1309.6392.pdf)



- Lime
  - https://arxiv.org/abs/1602.04938