Lecture 10:
Feature Importance estimation

Moscow, Russia

Outline

1. Feature importance estimation
2. Shap values

Feature importance estimation

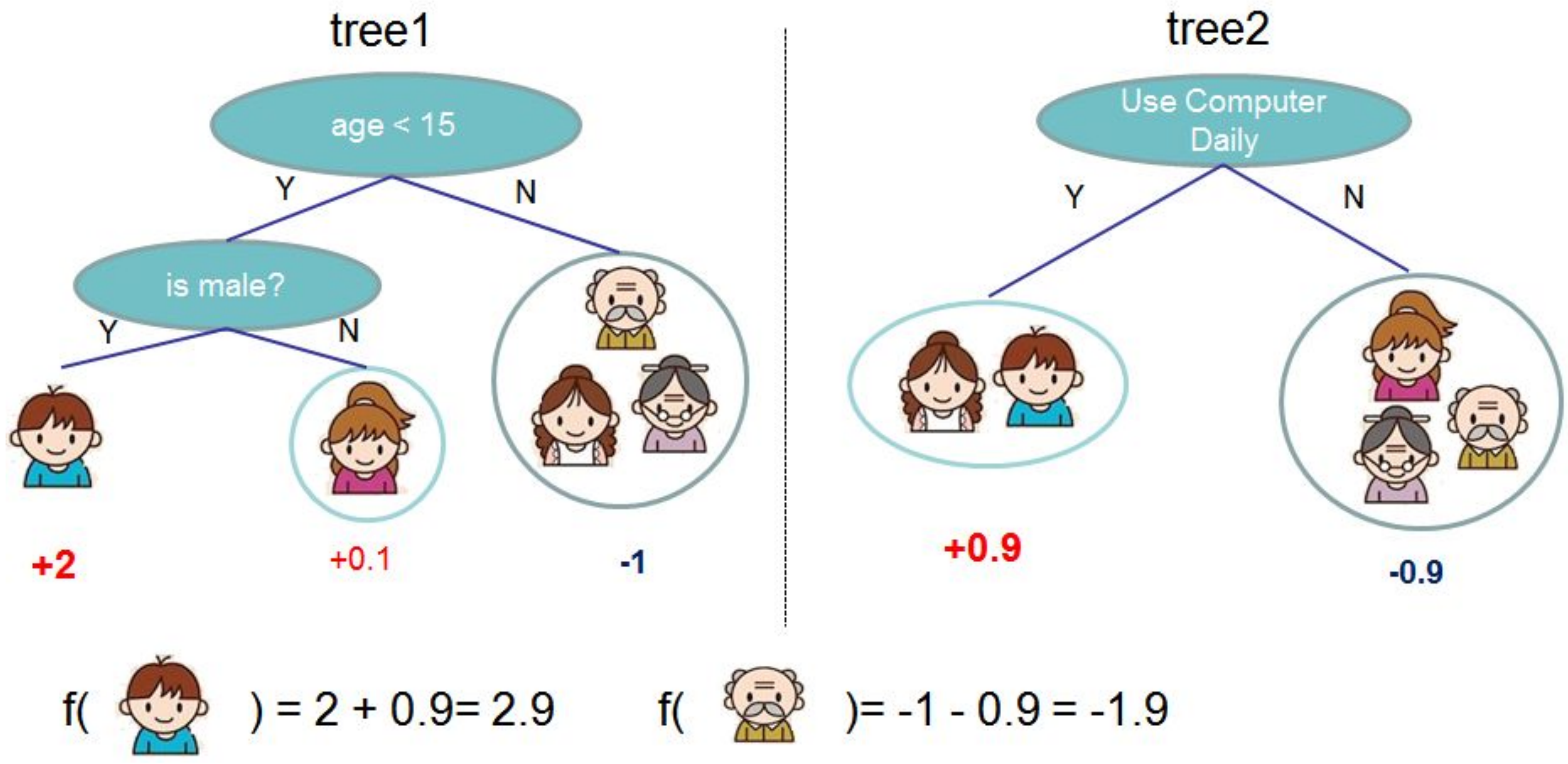

Image source: https://habr.com/ru/post/428213/

Feature importance estimation

1. Permutation importance
2. Partial Dependence Plots (PDP)
3. Tree specific:
    a. Gain
    b. Frequency (Split Count)
    c. Cover (weighted Split Count)
4. Shap

Permutation importance

Height at age 20 (cm)	Height at age 10 (cm)	...	Socks owned at age 10
182	155	...	20
175	147	...	10
...	...	...	...
156	142	...	8
153	130	...	24

Image source: https://www.kaggle.com/dansbecker/permutation-importance

Permutation importance

Height at age 20 (cm)	Height at age 10 (cm)	...	Socks owned at age 10
182	155	...	20
175	147	...	10
...	...	...	...
156	142	...	8
153	130	...	24

Train model

Observe changes caused by feature random permutations

Image source: https://www.kaggle.com/dansbecker/permutation-importance

Partial Dependence Plots

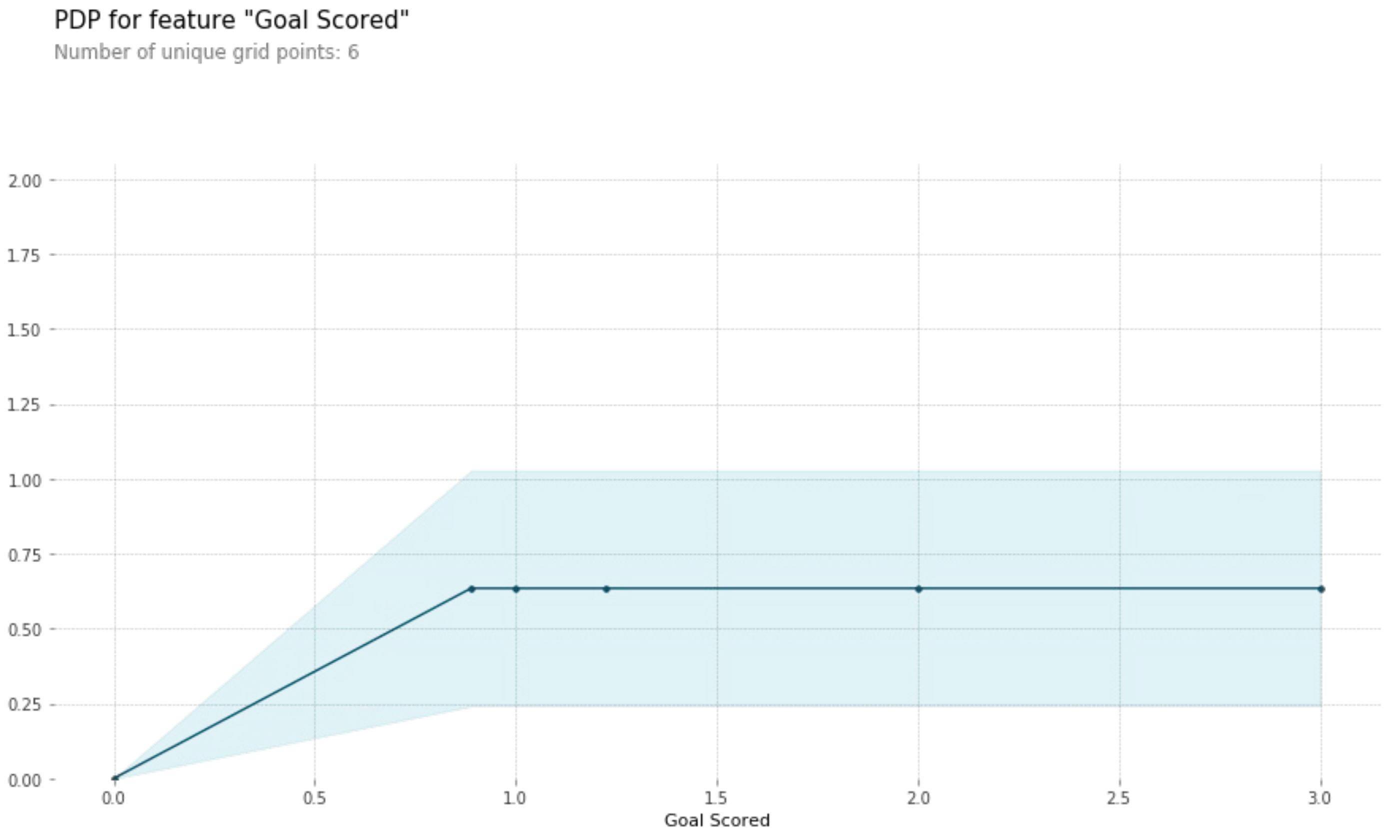

Image source: https://www.kaggle.com/dansbecker/partial-plots

Importance estimation problems

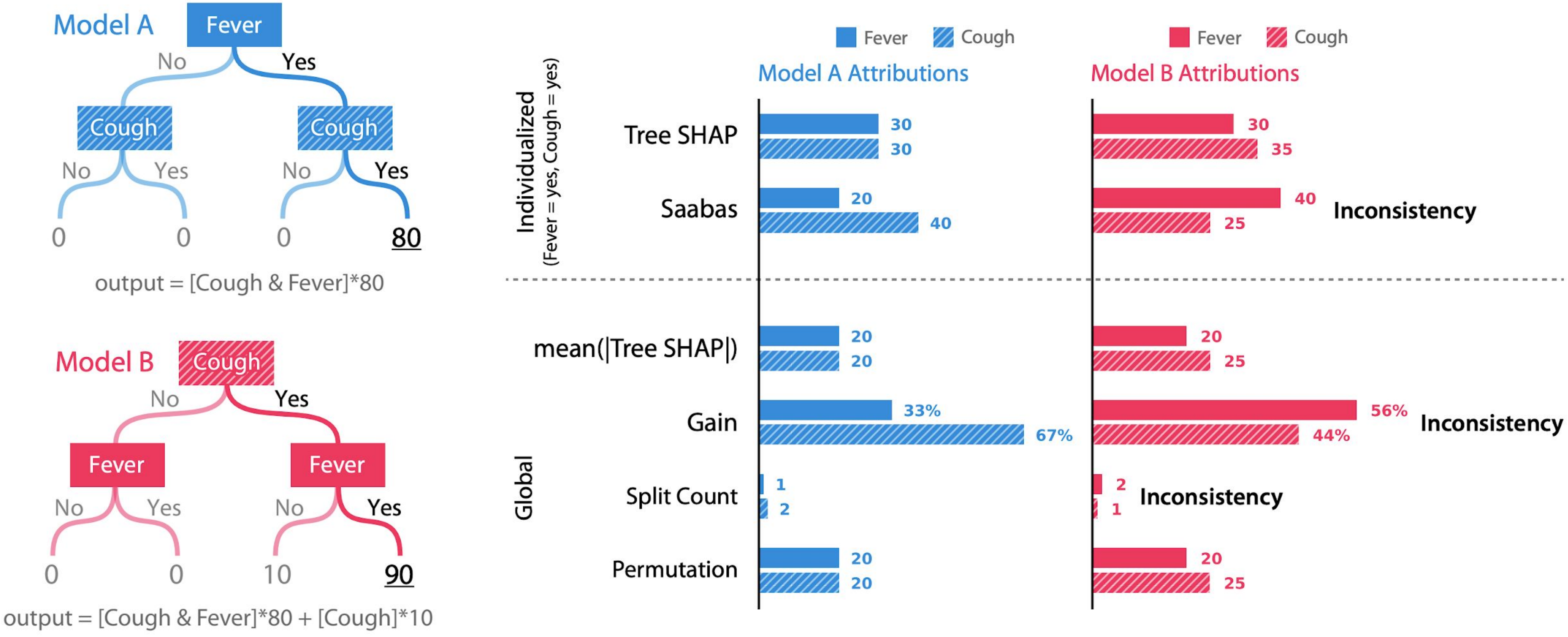

Image source: [“Consistent Individualized Feature Attribution for Tree Ensembles”](#) paper

Shap values

Consider *i-th* feature. Shap value will be

$$\phi_i(p) = \sum_{S \subseteq N/\{i\}} \frac{|S|!(n - |S| - 1)!}{n!}(p(S \cup \{i\}) - p(S))$$

where $p(S \cup \{i\})$ is model prediction on feature subset S with *i-th* feature added.

Shap values

Consider *i-th* feature. Shap value will be

$$\phi_i(p) = \sum_{S \subseteq N/\{i\}} \frac{|S|!(n - |S| - 1)!}{n!}(p(S \cup \{i\}) - p(S))$$

where $p(S \cup \{i\})$ is model prediction on feature subset S with *i-th* feature added.

SHAP values are the only consistent and locally accurate individualized feature attributions

See [Consistent Individualized Feature Attribution for Tree Ensembles] paper for more info

Outro

The rest is in notebook form