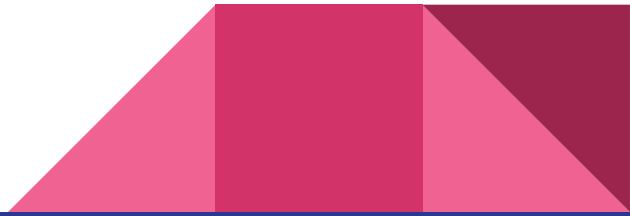
Examining Hoeffding Trees with SHAP

Seminar on Explainable and Fair Machine Learning

Outline

- Motivation: Why explainability?
- Additive Feature Attribution Methods
- The SHAP framework
  - Kernel SHAP
  - Tree SHAP
- Models
  - Decision Trees
  - Hoeffding Trees
- Experiments
- Conclusion



Why Explain?

- Build trust
- Improve model
- Gain insights into the problem
- Actionability
- Debugging ML pipelines
- Accountability

"gibbon"
99.3 % confidence

[3][4][5]

How to explain?

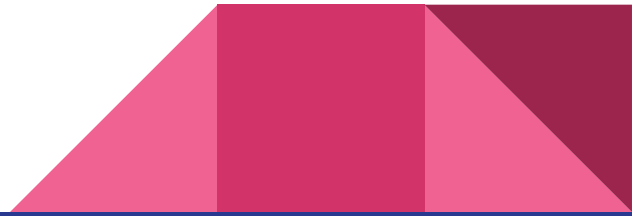
Additive Feature Attribution methods

- Explanation models
- Unifies many current methods
- Lime, Deep Lift, Layer-Wise Relevance Propagation, ...

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i$$

Guarantees?

- Local Accuracy
  - Explainer should make the same predictions as the original model
- Missingness
  - Missing features have no impact
- Consistency
  - If a feature becomes more important the attribution should not go down.



The only Solution!

Shapely Values

- comes from Game Theory
- proved to be the only solution satisfying all 3 properties
- average of each features marginal contribution considering all input feature permutations
- in general NP-Hard :(

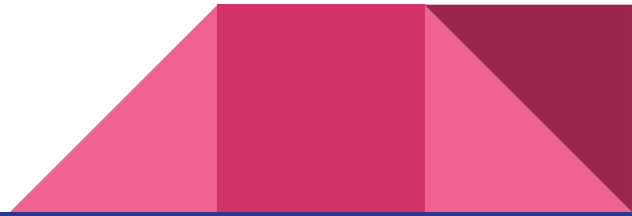
$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!}[f_x(z') - f_x(z' \setminus i)]$$

Lets compute them anyways!

Shapely Additive exPlanations SHAP

- Additive Feature Attribution method using Shapley values as features
- Algorithms:
  - Kernel SHAP
  - Linear SHAP
  - Max SHAP
  - Low-Order SHAP
  - Deep SHAP
  - Tree SHAP

Let’s have a look at two of these



The general Tool!

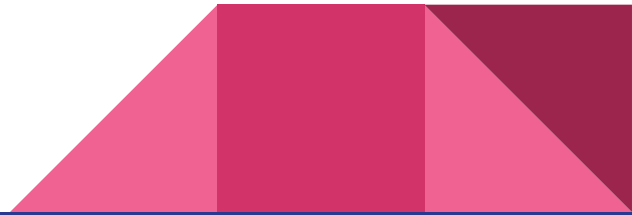
Kernel SHAP (Linear Lime + Shapley Values)(NIPS 2017)

- Chooses LIME loss and mapping so that it recovers Shapley values
- Model agnostic
- Approximates SHAP values
- Assumes model linearity and feature independence
- $O(M2^M)$
- Different from heuristic LIME

The superior algorithm!

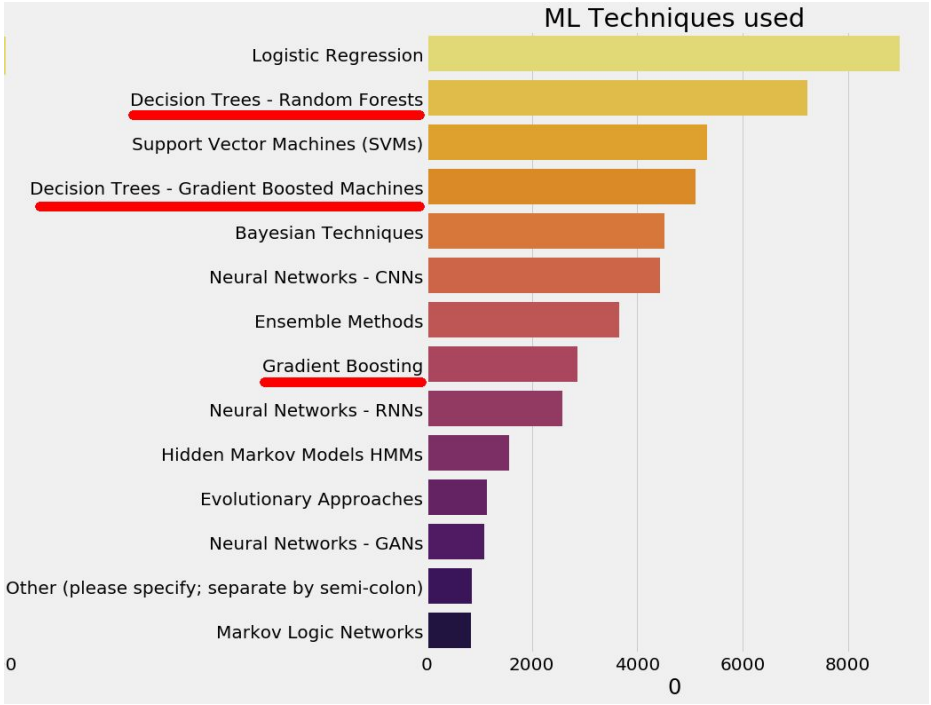
Tree SHAP (Nature Machine Intelligence 2020)

- specialized for tree models
- calculates **exact** SHAP values
- polynomial runtime: $O(TLD^2)$
- no new assumptions
- used in healthcare
- enables calculating shapley interaction values
- enables supervised clustering



Are Trees relevant?

Yes!

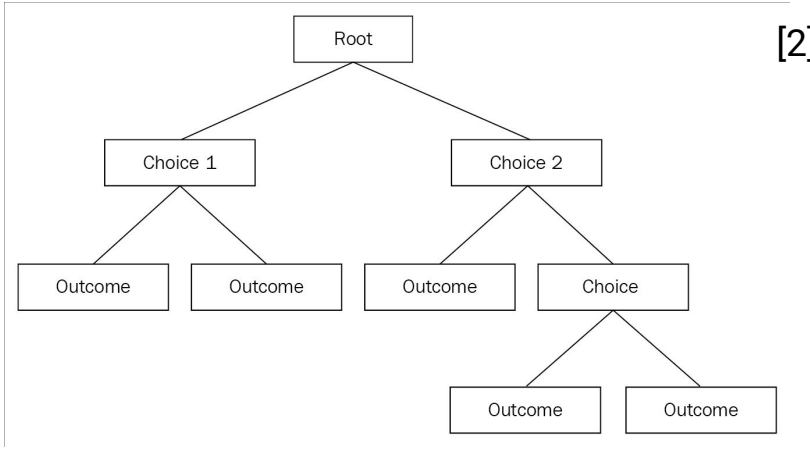


Kaggle Survey 2017 [1]

What’s a tree model?

Decision Trees:

- Start at root
- Test input on every node of tree
- Split based on heuristic and pass down
- Labels are assigned on the leaves
- Trained recursively



[2]

Hoeffding Tree??

Hoeffding Tree

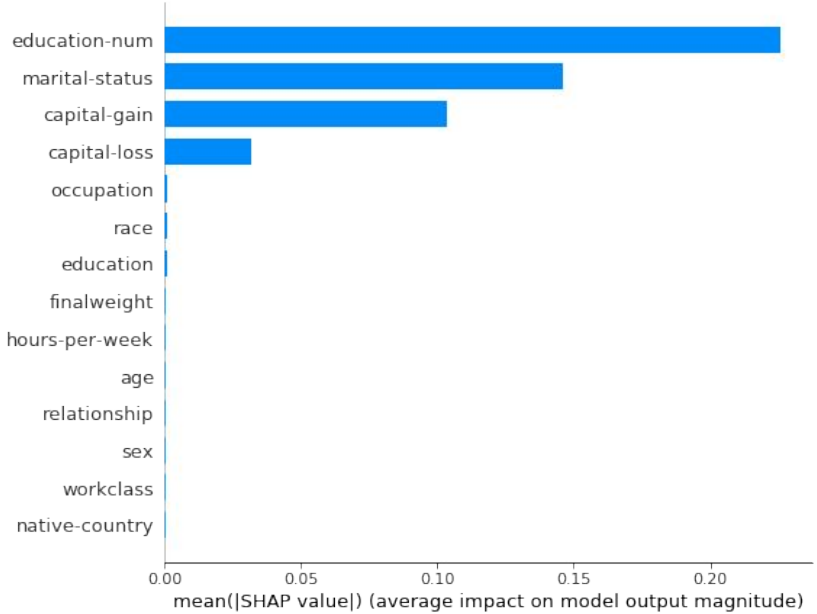
- Decision tree that can be trained on mini batches
- Uses Hoeffding bounds to determine when to split
- Used for mining high speed data streams
- Converges asymptotically towards batch learner

But how fast does it converge?

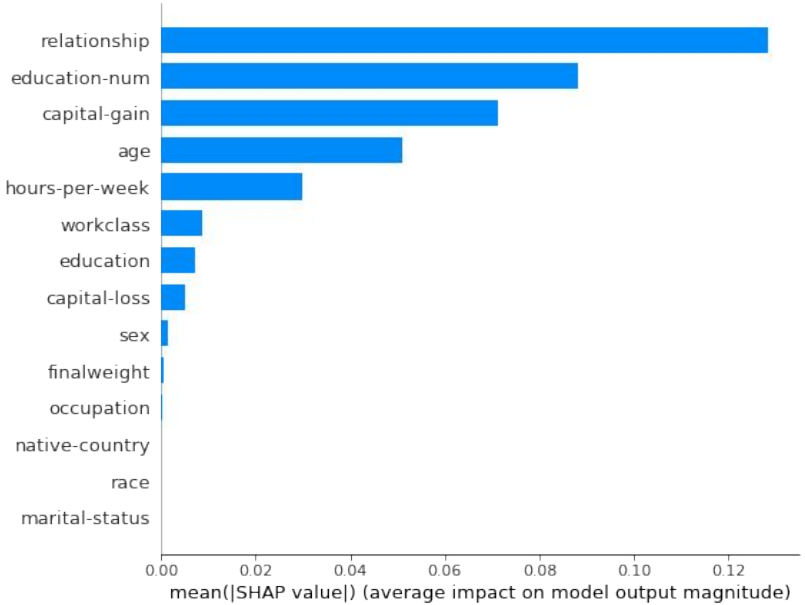
Experiments

- Trained decision and Hoeffding tree on UCI adults dataset
- Computed SHAP values of decision tree with Tree SHAP
- Tree SHAP not yet compatible with Hoeffding libraries :(
- -> back to Kernel SHAP
- Surprisingly observed no convergence.
- Reason unclear. Implementations? Too little data? Data drift?

Hoeffding Tree



Decision Tree



But can we trust Kernel SHAP?

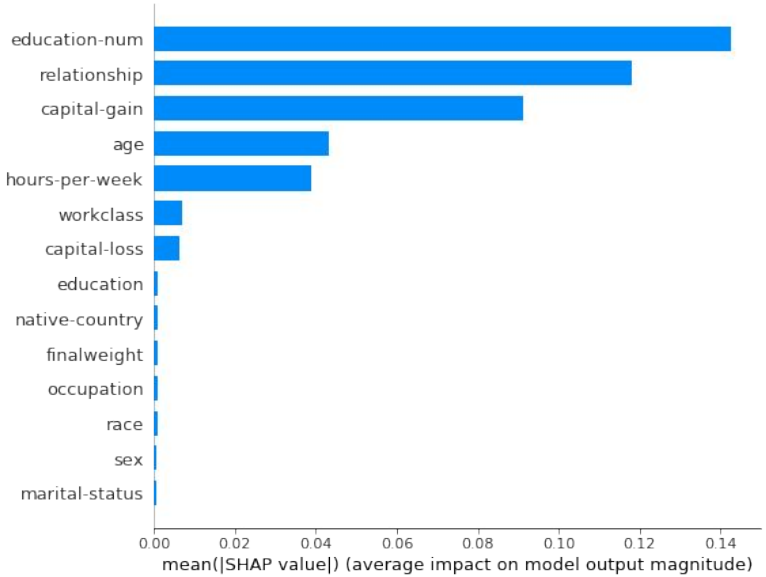
- Kernel SHAP only an approximation under assumptions
- We can compare it to Tree SHAP which is exact

-> Second experiment

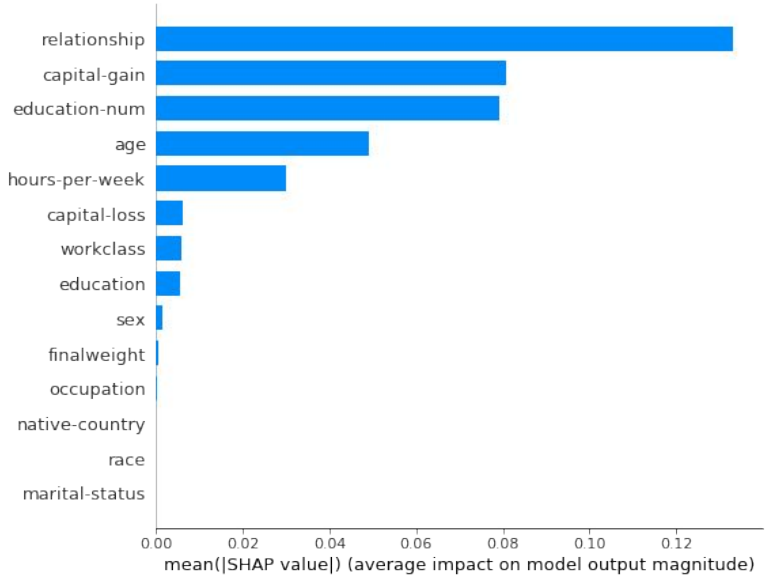
- Trained Decision tree again and explained it with both methods
- Results were reasonably close. Only 1 feature significantly off.

It’s not bad!

Kernel SHAP

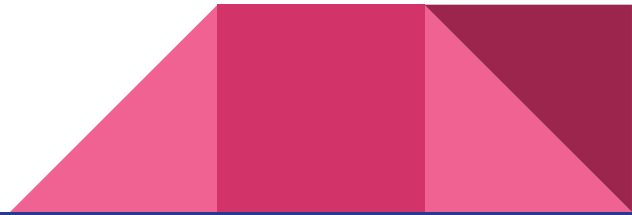


Tree Shap



Conclusion

- Additive Feature Attribution methods are a broad class of explanation models
- SHAP is the unique solution guaranteeing local accuracy, missingness and consistency
- Excellent algorithm for trees, approximations for others
- Did not observe convergence for Hoeffding tree.
  - Reason remains an open question
- Kernel SHAP seems to be a reasonable approximation overall
  - But: Individual features can be quite off.



References

[1]https://www.kaggle.com/ash316/novice-to-grandmaster

[2]https://static.packt-cdn.com/products/9781783553112/graphics/image_05_004.png

[3]https://i.kinja-img.com/gawker-media/image/upload/c_scale,f_auto,fl_progressive,pg_1,q_80,w_1600/yjud7idcynanyxaet43y.png

[4]https://medium.com/onfido-tech/adversarial-attacks-and-defences-for-convolutional-neural-networks-66915ece52e7

[5]https://techstartups.com/wp-content/uploads/2018/04/Art[illegible]l-Intelligen[illegible]-Healthcare-960x600.jpg

Thank you for your attention!