Implementation of the unbiased KernelSHAP

SHAP

- exact
- computation time exponential in number of features (assuming blackbox model)

KernelSHAP

- approximates SHAP values via linear regression
- finds **exact** solution to the **approximate** problem by sampling
- consistent, O(1/n) conv.
- *seems to be unbiased*
- *empirically found convergence formula and error estimates*

unbiased KernelSHAP

- *approximates SHAP values via linear regression*
- *finds **approximate** solution to the **exact** problem by sampling*
- *consistent, O(1/n) conv., slower than KernelSHAP*
- *provably unbiased*
- *higher variance than KernelSHAP*
- *derived convergence formula and error estimates*

Italic font indicates news from the paper. Moreover a better sampling strategy is proposed and tested to lower estimators' variances.
[link to the paper](link%20to%20the%20paper)