Multicollinearity Correction and Combined Feature Effect in Shapley Values

Indranil Basu, Data Science Manager, UHG | Subhadip Maji, Data Scientist, UHG

Abstract

Model interpretability is one of the most intriguing problems in most machine learning models, particularly for those that are mathematically sophisticated. Computing Shapley Values are one of the best approaches so far to find the importance of each feature in a model, at the instance (data point) level. In other words, Shapley values represent the importance of a feature for a particular instance or observation, especially for classification or regression problems. One of the well known limitations of Shapley values is that the estimation of Shapley values with the presence of multicollinearity among the features are not accurate as well as reliable. To address this problem, we present a unified framework to calculate accurate Shapley values with correlated features. To be more specific, we do an adjustment (matrix formulation) of the features while calculating independent Shapley values for the instances to make the features independent with each other. Our implementation of this method proves that our method is computationally efficient also, compared to the existing Shapley method.

MCC Shapley Values for Individual Features

Assume for a dataset the correlation of X_j with other features $X_1, X_2, ..., X_{j-1}$, $X_{j+1}, ..., X_m$ are $c_{j1}, c_{j2}, ..., c_{j(j-1)}, c_{j(j+1)}, ..., c_{jm}$ respectively. If we are interested in calculating the shapely value of X_j, we add one Adjustment Factor (AF_k) with X_k, where $k = 1, 2, ..., j-1, j+1, ..., m$, while we randomize (or remove) X_j in the Shapley value calculation process so that,

$$cor(X_j, X_k + AF_k) = 0 \tag{1}$$

Putting $AF_k = aX_j$ in the above equation and solving we get,

$$AF_k = -\frac{cov(X_j, X_k)}{var(X_j)} X_j \tag{2}$$

The reason for taking AF_k as only a function of X_j, because we want to remove the correlation effect of X_k only from X_j.

MCC Shapley Values for Combination of Two Features

Assume for a dataset the correlation of X_i and X_j with other features X_k, where $k = 1, 2, ..., m$ and $k \notin \{i, j\}$, are c_{ik} and c_{jk} respectively. If we are interested in calculating the shapely value of the combination of X_i and X_j, we add one Adjustment Factor (AF_k) with X_k, while we randomize (or remove) X_i and X_j all together in the Shapley value calculation process so that,

$$\begin{aligned} cor(X_i, X_k + AF_k) &= 0 \\ cor(X_j, X_k + AF_k) &= 0 \end{aligned} \tag{3}$$

Putting $AF_k = aX_i + bX_j$ in the above equation and solving we get,

$$\begin{aligned} a &= \frac{cov(X_i, X_k)var(X_j) - cov(X_j, X_k)cov(X_i, X_j)}{var(X_i)var(X_j) - (cov(X_i, X_j))^2} \\ b &= \frac{cov(X_j, X_k)var(X_i) - cov(X_i, X_k)cov(X_i, X_j)}{var(X_i)var(X_j) - (cov(X_i, X_j))^2} \end{aligned} \tag{4}$$

MCC Shapley Values for Combination of > 2 Features

$$\begin{aligned} cor(X_i, X_k + AF_k) &= 0 \\ cor(X_{i+1}, X_k + AF_k) &= 0 \\ cor(X_{i+2}, X_k + AF_k) &= 0 \\ &\vdots \\ cor(X_{i+p}, X_k + AF_k) &= 0 \end{aligned} \tag{5}$$

Putting,

$$AF_k = a_i X_i + a_{i+1} X_{i+1} + a_{i+2} X_{i+2} + ... + a_{i+p} X_{i+p} \tag{6}$$

in the equation 5 and writing the p equations in matrix form we get,

$$\begin{bmatrix} var(X_i) & cov(X_i, X_{i+1}) & cov(X_i, X_{i+2}) & ... & cov(X_i, X_{i+p}) \\ cov(X_{i+1}, X_i) & var(X_{i+1}) & cov(X_{i+1}, X_{i+2}) & ... & cov(X_{i+1}, X_{i+p}) \\ cov(X_{i+2}, X_i) & cov(X_{i+2}, X_{i+1}) & var(X_{i+2}) & ... & cov(X_{i+2}, X_{i+p}) \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ cov(X_{i+j}, X_i) & cov(X_{i+j}, X_{i+1}) & cov(X_{i+j}, X_{i+2}) & ... & cov(X_{i+j}, X_{i+p}) \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ cov(X_{i+p}, X_i) & cov(X_{i+p}, X_{i+1}) & cov(X_{i+p}, X_{i+2}) & ... & var(X_{i+p}) \end{bmatrix} \begin{bmatrix} X_i \\ X_{i+1} \\ X_{i+2} \\ ... \\ ... \\ ... \\ X_{i+j} \\ ... \\ ... \\ ... \\ X_{i+p} \end{bmatrix} = \begin{bmatrix} cov(X_i, X_k) \\ cov(X_{i+1}, X_k) \\ cov(X_{i+2}, X_k) \\ ... \\ ... \\ ... \\ cov(X_{i+j}, X_k) \\ ... \\ ... \\ ... \\ cov(X_{i+p}, X_k) \end{bmatrix} \tag{7}$$

MCC Shapley Values for Combination of > 2 Features

By Cramer's rule,

$$a_{i+j} = \frac{\det A_{j+1}}{\det A} \tag{8}$$

where, $A_{j+1} =$

$$\begin{bmatrix} var(X_i) & ... & cov(X_i, X_{i+j-1}) & cov(X_i, X_k) & ... & cov(X_i, X_{i+p}) \\ cov(X_{i+1}, X_i) & ... & cov(X_{i+1}, X_{i+j-1}) & cov(X_{i+1}, X_k) & ... & cov(X_{i+1}, X_{i+p}) \\ cov(X_{i+2}, X_i) & ... & cov(X_{i+2}, X_{i+j-1}) & cov(X_{i+2}, X_k) & ... & cov(X_{i+2}, X_{i+p}) \\ ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... \\ cov(X_{i+j}, X_i) & ... & cov(X_{i+j}, X_{i+j-1}) & cov(X_{i+j}, X_k) & ... & cov(X_{i+j}, X_{i+p}) \\ ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... \\ cov(X_{i+p}, X_i) & ... & cov(X_{i+p}, X_{i+j-1}) & cov(X_{i+p}, X_k) & ... & var(X_{i+p}) \end{bmatrix}$$

and, $A =$

$$\begin{bmatrix} var(X_i) & cov(X_i, X_{i+1}) & cov(X_i, X_{i+2}) & ... & cov(X_i, X_{i+p}) \\ cov(X_{i+1}, X_i) & var(X_{i+1}) & cov(X_{i+1}, X_{i+2}) & ... & cov(X_{i+1}, X_{i+p}) \\ cov(X_{i+2}, X_i) & cov(X_{i+2}, X_{i+1}) & var(X_{i+2}) & ... & cov(X_{i+2}, X_{i+p}) \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ cov(X_{i+j}, X_i) & cov(X_{i+j}, X_{i+1}) & cov(X_{i+j}, X_{i+2}) & ... & cov(X_{i+j}, X_{i+p}) \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ cov(X_{i+p}, X_i) & cov(X_{i+p}, X_{i+1}) & cov(X_{i+p}, X_{i+2}) & ... & var(X_{i+p}) \end{bmatrix}$$

MCC Shapley Values for Combination of > 2 Features

Algorithm

Algorithm 1 Estimation of MCC Shapley Values

Output: MCC Shapley value for the value of the j-th feature, $x_j^{(i)}$.
Required: Number of iterations M, instance of interest $x^{(i)}$, feature index j, data matrix $\mathbf{X}$, and machine learning model f
for $m \leftarrow 1, M$ **do**
  Draw random instance $x^{(r)}$ from the data matrix X
  Choose a random permutation o of the feature values
  Order instance $x^{(i)}$: $x_{[o]}^{(i)} \leftarrow (x_{(1)}^{(i)}, ..., x_{(j)}^{(i)}, ..., x_{(m)}^{(i)})$
  Order instance $x^{(r)}$: $x_{[o]}^{(r)} \leftarrow (x_{(1)}^{(r)}, ..., x_{(j)}^{(r)}, ..., x_{(m)}^{(r)})$
  Construct two new instances adding adjustment factors to the feature values in coalitions i.e features belongs to instance $x^{(i)}$
    • With feature j: $x_{+j} \leftarrow (x_{(1)}^{(i)} + AF_{(1)}, ..., x_{(j-1)}^{(i)} + AF_{(j-1)}, x_{(j)}^{(i)}, x_{(j+1)}^{(r)}, ..., x_{(m)}^{(r)})$
    • Without feature j: $x_{-j} \leftarrow (x_{(1)}^{(i)} + AF_{(1)}, ..., x_{(j-1)}^{(i)} + AF_{(j-1)}, x_{(j)}^{(r)}, x_{(j+1)}^{(r)}, ..., x_{(m)}^{(r)})$
  Compute marginal contribution: $\phi_j^m \leftarrow \hat{f}(x_{+j}) - \hat{f}(x_{-j})$
Compute MCC Shapley value as the average: $\phi_j(x) \leftarrow \frac{1}{M} \sum_{m=1}^{M} \phi_j^m$

Results on House Prices Dataset

Results of MCC Shapley Values for Individual Features

Table 1. Shapley Values with and without Multi-collinearity Correction for a randomly picked data point for `MiscVal` Feature. These values are created for a Monte-Carlo simulation with 10000 iterations.

Model	Without Presence of Artificially Created Variable		With Presence of Artificially Created Variable			
	NMCC-SV of MiscVal	MCC-SV of MiscVal	NMCC-SV of MiscVal	NMCC-SV of MiscVal_corr	MCC-SV of MiscVal	MCC-SV of MiscVal_corr
Decision Tree	-261.4 ± 5.3	-260.2 ± 5.3	-129.3 ± 5.1	-128.3 ± 4.8	-263.1 ± 5.4	-262.9 ± 4.7
Random Forest	-210.9 ± 1.4	-209.8 ± 1.4	-107.8 ± 1.1	-102.4 ± 1.2	-213.5 ± 0.9	-210.2 ± 1.3
Gradient Boosting	-273.2 ± 1.1	-272.1 ± 1.1	-138.1 ± 1.4	-137.3 ± 1.2	-276.3 ± 1.8	-275.2 ± 1.0
Extreme Gradient Boosting	-265.9 ± 1.2	-263.8 ± 1.2	-135.2 ± 1.2	-131.7 ± 1.3	-265.0 ± 0.9	-264.2 ± 1.1
Support Vector Regression	-112.6 ± 0.4	-113.9 ± 0.4	-53.4 ± 0.3	-55.6 ± 0.1	-111.0 ± 0.1	-115.7 ± 0.2

From Table it is seen that with the presence of **MiscVal_corr** the NMCC Shapley values have sliced to half of the NMCC Shapley values without the presence of **MiscVal_corr** for all the models.

Table 2. Shapley Values with and without Multi-collinearity Correction for a randomly picked data point for `1stFlrSF` Feature. These values are created for a Monte-Carlo simulation with 10000 iterations.

Model	Without Presence of TotalBsmtSF		With Presence of all of 331 features	
	NMCC-SV of 1stFlrSF	MCC-SV of 1stFlrSF	NMCC-SV of 1stFlrSF	MCC-SV of 1stFlrSF
Decision Tree	2721.1 ± 5.1	2832.2 ± 5.0	1932.8 ± 5.6	2841.7 ± 5.2
Random Forest	2502.3 ± 1.1	2715.4 ± 1.6	1781.3 ± 1.2	2700.4 ± 1.5
Gradient Boosting	2657.7 ± 0.9	2919.1 ± 1.1	1699.1 ± 0.9	2933.4 ± 1.0
Extreme Gradient Boosting	3356.2 ± 1.4	3612.6 ± 0.9	2134.5 ± 1.5	3597.8 ± 1.1
Support Vector Regression	1745.8 ± 0.5	1983.7 ± 0.4	1244.9 ± 0.7	1998.4 ± 0.8

we see that the NMCC Shapley values have reduced down a lot, but the reduction is not roughly 50% like scenario 1, because here the correlation between **TotalBsmtSF** and **1stFlrSF** is not 1.

Results on House Prices Dataset

Results of MCC Shapley Values for Combination of two Features

Table 3. Shapley Values with and without Multi-collinearity Correction for a randomly picked data point for the combination of `MiscVal` and `3SsnPorch` features. These values are created for a Monte-Carlo simulation with 10000 iterations.

Model	Without Presence of Artificially Created Variable		With Presence of Artificially Created Variable	
	NMCC-SV of the combination of MiscVal and 3SsnPorch	MCC-SV of the combination of MiscVal and 3SsnPorch	NMCC-SV of the combination of MiscVal and 3SsnPorch	MCC-SV of the combination of MiscVal and 3SsnPorch
Decision Tree	417.2 ± 4.9	415.6 ± 5.0	211.4 ± 5.1	419.6 ± 4.8
Random Forest	323.6 ± 1.9	327.8 ± 1.8	163.8 ± 1.9	330.9 ± 2.0
Gradient Boosting	374.3 ± 1.4	376.1 ± 1.7	185.0 ± 1.3	370.7 ± 1.5
Extreme Gradient Boosting	289.5 ± 1.1	292.5 ± 1.4	149.7 ± 1.7	289.1 ± 1.6
Support Vector Regression	134.5 ± 0.7	145.3 ± 0.8	67.3 ± 0.9	149.2 ± 0.6

correction the MCC Shapley values increase with respect to their NMCC Shapley values counter-parts.

From the result, it is seen that the NMCC Shapley values sliced to half due to the presence of the perfectly correlated features, but the multi-collinearity correction factor helped those features to get back to the actual values i.e. MCC Shapley values. Also, as **MiscVal** and **3SsnPorch** are almost uncorrelated with other features, NMCC and MCC Shapley values are almost the same without the presence of artificially created variable across all the models.

Table 4. Shapley Values with and without Multi-collinearity Correction for a randomly picked data point for the combination of `1stFlrSF` and `2ndFlrSF` features. These values are created for a Monte-Carlo simulation with 10000 iterations.

Model	NMCC-SV of combination of 1stFlrSF and 2ndFlrSF	MCC-SV of combination of 1stFlrSF and 2ndFlrSF
Decision Tree	3321.5 ± 3.2	4610.3 ± 3.3
Random Forest	2895.4 ± 1.4	3767.6 ± 1.6
Gradient Boosting	3006.7 ± 1.3	4209.2 ± 1.3
Extreme Gradient Boosting	3209.1 ± 0.9	4479.9 ± 1.0
Support Vector Regression	3877.0 ± 0.7	5003.6 ± 0.9

Execution Time Assessment

We performed one additional experiment where we compared the execution time between the MCC and NMCC Shapley value calculation. This experiment is done in a machine with 2.6 GHz Intel Core i7 and 8 GB available RAM. Following Table shows the comparison in execution time of the Shapley Values with and without multi-collinearity correction. This result is produced with a Random Forest model which is trained on default parameters and the number of iterations of Monte-Carlo simulation is 10,000. From the table, it is seen that introduction of the correlation adjustment factor has almost no effect on the execution time for calculating Shapley values.

Feature Size	NMCC-SV(sec)	MCC-SV(sec)
≈ 10	0.3 ± 0.01	0.3 ± 0.03
≈ 100	0.9 ± 0.02	1.0 ± 0.03
≈ 1000	4.7 ± 0.01	4.9 ± 0.01
≈ 10000	17.3 ± 0.01	18.1 ± 0.01