

ITEA automatic report

ITEA_summarizer

Friday 2nd July, 2021, 20:52

Automatic report created by *ITEA_summarizer* package. This report makes usage of several methods to automatically inspect and explain the final expression found in the evolutionary process performed by the ITEA algorithm.

Descriptive statistics of the data

Reporting descriptive statistics for 5 (from a total of 8) features contained on the training data. The features were selected based on the absolute final importance.

	AveBedrms	MedInc	AveOccup	AveRooms	Latitude
count	13828.000000	13828.000000	13828.000000	13828.000000	13828.000000
mean	1.097533	3.876745	3.128660	5.436556	35.651238
std	0.445688	1.903102	12.646130	2.449446	2.134064
min	0.333333	0.499900	0.692308	0.888889	32.550000
25%	1.006623	2.568575	2.432189	4.459802	33.940000
50%	1.049552	3.538750	2.819702	5.232422	34.270000
75%	1.100283	4.756600	3.282093	6.058566	37.720000
max	25.636364	15.000100	1243.333333	141.909091	41.950000

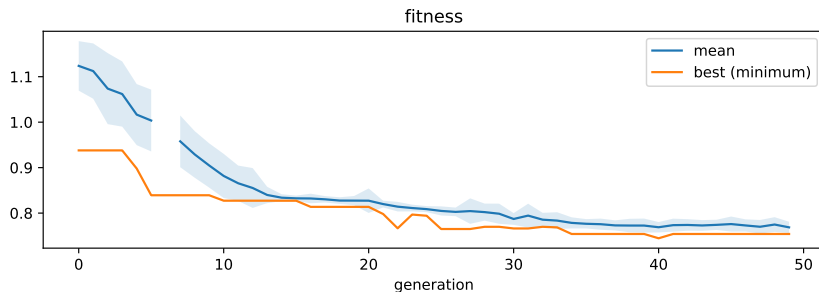
Algorithm Hyper-parameters

The following hyperparameters were used to execute the algorithm. If the `random_state` parameter was set to an integer value (or a `numpy.randomState` instance was given), then it is possible to repeat the exact execution by using the same training data and the parameters listed below.

```
expolim : (0, 2)
gens : 50
max_terms : 5
popsize : 50
random_state : 42
simplify_method : simplify_by_var
verbose : 10
tfuncs : [log, sqrt.abs, id, sin, cos, exp]
```

Evolution convergence

The algorithm took 127.554 seconds to completely run. Below are the plots for the average fitness of the population and the best individual fitness for each generation.



Best expression

The best expression corresponds to the expression with the best fitness on the last generation before the evolution ends. Not necessarily it will be the simplest or the global optimum expression of the evolution. The final expression is a regressor with a fitness of 0.75408, and the number of IT terms is 5. Below is a LaTeX representation of the expression:

$$\begin{aligned}
 ITEXpr = & \underbrace{\beta_0 \cdot \log(\text{MedInc}^2 \cdot \text{AveBedrms} \cdot \text{Longitude}^2)}_{\text{term 0}} \\
 & + \underbrace{\beta_1 \cdot \log(\text{MedInc} \cdot \text{HouseAge} \cdot \text{AveRooms} \cdot \text{AveOccup}^2 \cdot \text{Longitude}^2)}_{\text{term 1}} \\
 & + \underbrace{\beta_2 \cdot \log(\text{HouseAge} \cdot \text{AveRooms} \cdot \text{AveBedrms} \cdot \text{AveOccup}^2 \cdot \text{Latitude} \cdot \text{Longitude}^2)}_{\text{term 2}} \\
 & + \underbrace{\beta_3 \cdot \log(\text{HouseAge}^2 \cdot \text{AveBedrms} \cdot \text{AveOccup}^2 \cdot \text{Latitude}^2 \cdot \text{Longitude}^2)}_{\text{term 3}} \\
 & + \underbrace{\beta_4 \cdot \log(\text{MedInc})}_{\text{term 4}} + I_0
 \end{aligned} \tag{1}$$

Best expression metrics

On the next page is reported a table containing the coefficients for the previous expression, as well as some metrics calculated for each term individually:

- **coef:** coefficient of each term (or coefficients, if the itexpr is an instance of IT-Expr_classifier);
- **coef stderr:** the standard error of the coefficients;
- **disentang.:** mean pairwise disentanglement between each term when compared with the others;
- **M.I.:** mean continuous mutual information between each term when compared with the others;
- **pred. var.:** variance of the predicted outcomes for each term when predicting the training data.

	coef	func	coef stderr	disentang.	M.I.	pred. var.
term 0	9.924	log	0.238	0.323	0.748	85.569
term 1	7.982	log	0.223	0.426	0.407	57.724
term 2	-9.092	log	0.222	0.361	0.439	55.374
term 3	0.702	log	0.013	0.312	0.379	0.778
term 4	-25.846	log	0.697	0.324	0.754	147.950
term 5	-62.377	intercept	1.501	0.000	0.000	0.000

Partial derivatives

$$\frac{\partial}{\partial MedInc} ITEXpr \quad (2)$$

$$= \underbrace{2\beta_0 \cdot \log'(MedInc^2 \cdot AveBedrms \cdot Longitude^2)}_{\text{term 0}} (MedInc \cdot AveBedrms \cdot Longitude^2) \\ + \underbrace{1\beta_1 \cdot \log'(MedInc \cdot HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2)}_{\text{term 1}} (HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2) \\ + \underbrace{1\beta_4 \cdot \log'(MedInc)}_{\text{term 4}} ()$$

$$\frac{\partial}{\partial HouseAge} ITEXpr \quad (3)$$

$$= \underbrace{1\beta_1 \cdot \log'(MedInc \cdot HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2)}_{\text{term 1}} (MedInc \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2) \\ + \underbrace{1\beta_2 \cdot \log'(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} (AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2) \\ + \underbrace{2\beta_3 \cdot \log'(HouseAge^2 \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} (HouseAge \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)$$

$$\begin{aligned}
& \frac{\partial}{\partial AveRooms} IExpr \\
&= \underbrace{1\beta_1 \cdot \log'(MedInc \cdot HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2)}_{\text{term 1}} (MedInc \cdot HouseAge \cdot AveOccup^2 \cdot Longitude^2) \\
&+ \underbrace{1\beta_2 \cdot \log'(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} (HouseAge \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)
\end{aligned} \tag{4}$$

$$\begin{aligned}
& \frac{\partial}{\partial AveBedrms} IExpr = \underbrace{1\beta_0 \cdot \log'(MedInc^2 \cdot AveBedrms \cdot Longitude^2)}_{\text{term 0}} (MedInc^2 \cdot Longitude^2) \\
&+ \underbrace{1\beta_2 \cdot \log'(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} (HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2) \\
&+ \underbrace{1\beta_3 \cdot \log'(HouseAge^2 \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} (HouseAge^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)
\end{aligned} \tag{5}$$

$$\frac{\partial}{\partial Population} IExpr = 0.0 \tag{6}$$

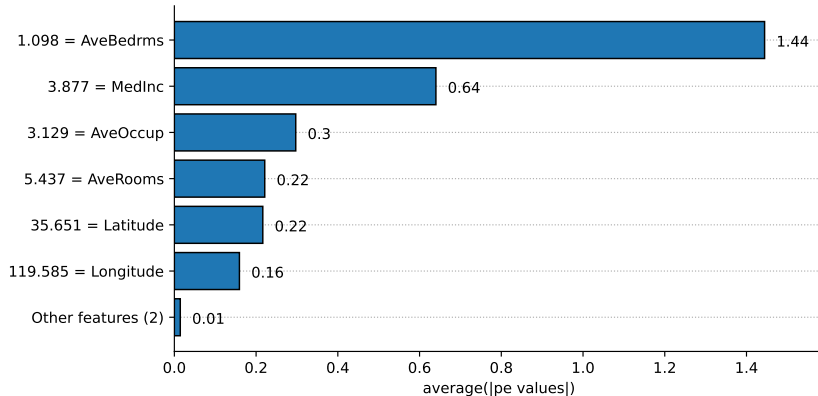
$$\begin{aligned}
& \frac{\partial}{\partial AveOccup} IExpr \tag{7} \\
&= 2\beta_1 \cdot \log'(MedInc \cdot HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2) \underbrace{(MedInc \cdot HouseAge \cdot AveRooms \cdot AveOccup^2 \cdot Longitude^2)}_{\text{term 1}} \\
&\quad + 2\beta_2 \cdot \log'(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2) \underbrace{(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} \\
&\quad + 2\beta_3 \cdot \log'(HouseAge^2 \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2) \underbrace{(HouseAge^2 \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}}
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial}{\partial Latitude} IExpr \tag{8} \\
&= 1\beta_2 \cdot \log'(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2) \underbrace{(HouseAge \cdot AveRooms \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} \\
&\quad + 2\beta_3 \cdot \log'(HouseAge^2 \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2) \underbrace{(HouseAge^2 \cdot AveBedrms \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \text{Longitude}} \text{ITExpr} = & \underbrace{2\beta_0 \cdot \log'(\text{MedInc}^2 \cdot \text{AveBedrms} \cdot \text{Longitude}^2)}_{\text{term 0}} (\text{MedInc}^2 \cdot \text{AveBedrms} \cdot \text{Longitude}^2) \\
& + \underbrace{2\beta_1 \cdot \log'(\text{MedInc} \cdot \text{HouseAge} \cdot \text{AveRooms} \cdot \text{AveOccup}^2 \cdot \text{Longitude}^2)}_{\text{term 1}} (\text{MedInc} \cdot \text{HouseAge} \cdot \text{AveRooms} \cdot \text{AveOccup}^2 \cdot \text{Longitude}^2) \\
& + \underbrace{2\beta_2 \cdot \log'(\text{HouseAge} \cdot \text{AveRooms} \cdot \text{AveBedrms} \cdot \text{AveOccup}^2 \cdot \text{Latitude} \cdot \text{Longitude}^2)}_{\text{term 2}} (\text{HouseAge} \cdot \text{AveRooms} \cdot \text{AveBedrms} \cdot \text{AveOccup}^2 \cdot \text{Latitude} \cdot \text{Longitude}^2) \\
& + \underbrace{2\beta_3 \cdot \log'(\text{HouseAge}^2 \cdot \text{AveBedrms} \cdot \text{AveOccup}^2 \cdot \text{Latitude}^2 \cdot \text{Longitude}^2)}_{\text{term 3}} (\text{HouseAge}^2 \cdot \text{AveBedrms} \cdot \text{AveOccup}^2 \cdot \text{Latitude}^2 \cdot \text{Longitude}^2)
\end{aligned}$$

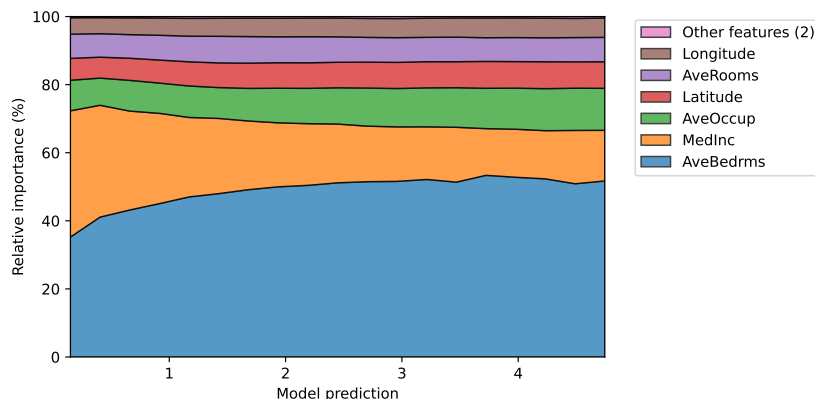
Average partial Effects

Feature importances with Average Partial Effects. This method attributes the importance to the i -th variable by calculating the average of the partial derivative w.r.t. i , evaluated for all data in the training set.



Normalized partial Effects

Feature importances with Normalized Partial Effects. To create this plot, first, the output interval is discretized. Then, for each interval, the partial effect of all samples in the training set that results in a prediction within the interval are calculated. Finally, they are normalized in order to make the total contribution by 100%.



Partial Effects at the Means

Partial Effects plots created by fixing the co-variables at the means and evaluating the model's output when only one variable changes. For simplicity, at most 5 variables are selected to create the plot (the 5 most important variables considering their Average Partial Effects).

