

# ITEA automatic report

*ITEA\_summarizer*

Friday 26<sup>th</sup> November, 2021, 16:48

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	Longitude	Latitude	AveOccup
count	13828.000000	13828.000000	13828.000000	13828.000000	13828.000000
mean	1.097533	3.876745	-119.585098	35.651238	3.128660
std	0.445688	1.903102	2.005127	2.134064	12.646130
min	0.333333	0.499900	-124.350000	32.550000	0.692308
25%	1.006623	2.568575	-121.810000	33.940000	2.432189
50%	1.049552	3.538750	-118.510000	34.270000	2.819702
75%	1.100283	4.756600	-118.010000	37.720000	3.282093
max	25.636364	15.000100	-114.310000	41.950000	1243.333333

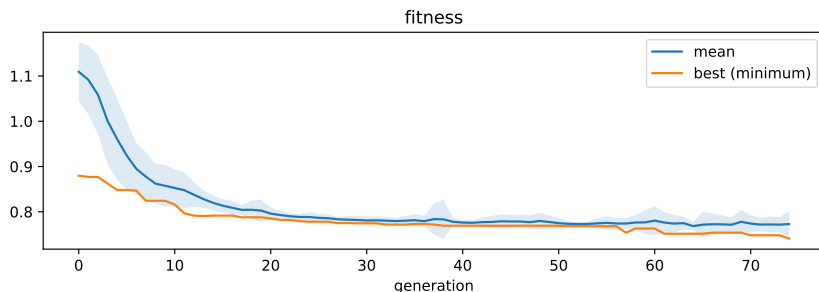
## Algorithm Hyper-parameters

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)
fitness_f : None
gens : 75
max_terms : 5
popsize : 75
random_state : 42
simplify_method : simplify_by_var
verbose : 10
tfuncs : [log, sqrt.abs, id, sin, cos, exp]
```

## Evolution convergence

The algorithm took 322.036 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 0.741, and the number of IT terms is 5. Below is a LaTeX representation of the expression:

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

## 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	-15.231	log	0.331	0.610	1.399	214.205
term 1	1.323	log	0.02	0.377	0.499	1.521
term 2	37.981	log	0.816	0.610	1.486	1341.802
term 3	-0.399	log	0.014	0.376	0.508	0.301
term 4	-11.439	log	0.273	0.609	1.439	487.150
term 5	-202.481	intercept	4.269	0.000	0.000	0.000

---

## Partial derivatives

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

$$\frac{\partial}{\partial HouseAge} ITEpr = 0.0 \tag{3}$$

$$\begin{aligned}
 & \frac{\partial}{\partial AveRooms} ITEpr \\
 = & \underbrace{2\beta_3 \cdot \log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} (MedInc^2 \cdot AveRooms \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)
 \end{aligned} \tag{4}$$



$$\begin{aligned} \frac{\partial}{\partial AveBedrms} ITEpr = & \underbrace{1\beta_1 \cdot \log'(MedInc^2 \cdot AveBedrms \cdot Longitude^2)}_{\text{term 1}} (MedInc^2 \cdot Longitude^2) \\ & + \underbrace{1\beta_3 \cdot \log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} (MedInc^2 \cdot AveRooms^2 \cdot \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{\partial}{\partial Population} ITEpr \\ = & \underbrace{1\beta_0 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude^2)}_{\text{term 0}} (MedInc \cdot AveOccup \cdot Latitude^2) \\ & + \underbrace{1\beta_2 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} (MedInc \cdot AveOccup \cdot Latitude \cdot Longitude^2) \\ & + \underbrace{2\beta_4 \cdot \log'(MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 4}} (MedInc^2 \cdot Population \cdot AveOccup^2 \cdot \end{aligned} \quad (6)$$

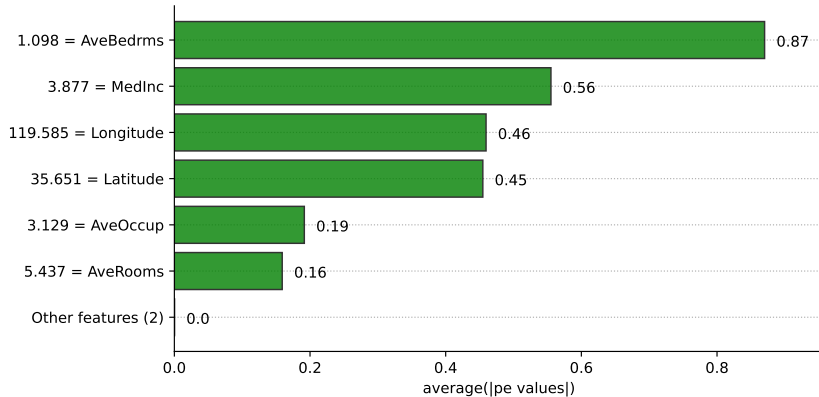
$$\begin{aligned}
& \frac{\partial}{\partial AveOccup} ITEXpr \\
&= \underbrace{1\beta_0 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude^2)(MedInc \cdot Population \cdot Latitude^2)}_{\text{term 0}} \\
&+ \underbrace{1\beta_2 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude \cdot Longitude^2)(MedInc \cdot Population \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} \\
&+ \underbrace{1\beta_3 \cdot \log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} \\
&+ \underbrace{2\beta_4 \cdot \log'(MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 4}}
\end{aligned} \tag{7}$$

$$\begin{aligned}
& \frac{\partial}{\partial Latitude} ITEXpr \\
&= \underbrace{2\beta_0 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude^2)(MedInc \cdot Population \cdot AveOccup \cdot Latitude^2)}_{\text{term 0}} \\
&+ \underbrace{1\beta_2 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude \cdot Longitude^2)(MedInc \cdot Population \cdot AveOccup \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} \\
&+ \underbrace{2\beta_3 \cdot \log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} \\
&+ \underbrace{2\beta_4 \cdot \log'(MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 4}}
\end{aligned} \tag{8}$$

$$\begin{aligned}
\frac{\partial}{\partial Longitude} ITE_{\text{expr}} = & \underbrace{2\beta_1 \cdot \log'(MedInc^2 \cdot AveBedrms \cdot Longitude^2)}_{\text{term 1}} (MedInc^2 \cdot AveBedrms \cdot Longitude^2) \\
& + \underbrace{2\beta_2 \cdot \log'(MedInc \cdot Population \cdot AveOccup \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} (MedInc \cdot Population \cdot AveOccup \cdot Latitude \cdot Longitude^2) \\
& + \underbrace{2\beta_3 \cdot \log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 3}} (MedInc^2 \cdot AveRooms^2 \cdot AveBedrms \cdot AveOccup \cdot Latitude^2 \cdot Longitude^2) \\
& + \underbrace{2\beta_4 \cdot \log'(MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 4}} (MedInc^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)
\end{aligned}$$

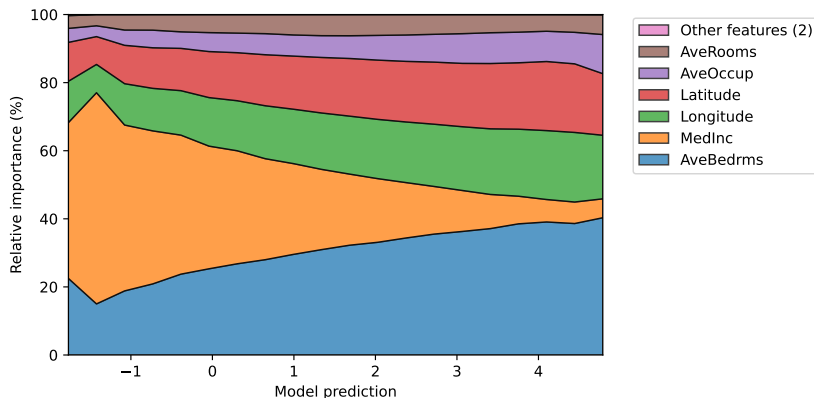
## Global importances with *Average partial Effects*

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).

