

Package ‘RcausalEGM’

January 20, 2025

Type Package

Title A General Causal Inference Framework by Encoding Generative Modeling

Version 0.3.3

Date 2023-03-22

Maintainer Qiao Liu <liuqiao@stanford.edu>

Description CausalEGM is a general causal inference framework for estimating causal effects by encoding generative modeling, which can be applied in both discrete and continuous treatment settings. A description of the methods is given in Liu (2022) <[arXiv:2212.05925](#)>.

License MIT + file LICENSE

Encoding UTF-8

Imports reticulate

URL <https://github.com/SUwonglab/CausalEGM>

BugReports <https://github.com/SUwonglab/CausalEGM/issues>

Repository CRAN

Depends R(>= 3.6.0)

RoxygenNote 7.2.3

Suggests rmarkdown, knitr, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Qiao Liu [aut, cre],
Wing Wong [aut],
Balasubramanian Narasimhan [ctb]

Date/Publication 2023-03-28 12:00:11 UTC

Contents

causalegm	2
get_est	4
install_causalegm	5

Index**6**

causalegm	<i>Main function for estimating causal effect in either binary or continuous treatment settings.</i>
-----------	--

Description

This function takes observation data (x,y,v) as input, and estimate the ATE/ITE/ADRF.

Usage

```
causalegm(  
  x,  
  y,  
  v,  
  z_dims = c(3, 3, 6, 6),  
  output_dir = ".",  
  dataset = "myData",  
  lr = 2e-04,  
  bs = 32,  
  alpha = 1,  
  beta = 1,  
  gamma = 10,  
  g_d_freq = 5,  
  g_units = c(64, 64, 64, 64, 64),  
  e_units = c(64, 64, 64, 64, 64),  
  f_units = c(64, 32, 8),  
  h_units = c(64, 32, 8),  
  dv_units = c(64, 32, 8),  
  dz_units = c(64, 32, 8),  
  save_model = FALSE,  
  save_res = FALSE,  
  binary_treatment = TRUE,  
  use_z_rec = TRUE,  
  use_v_gan = TRUE,  
  random_seed = 123,  
  n_iter = 30000,  
  normalize = FALSE,  
  x_min = NULL,  
  x_max = NULL  
)
```

Arguments

x	is the treatment variable, one-dimensional array with size n.
y	is the potential outcome, one-dimensional array with size n.

<code>v</code>	is the covariates, two-dimensional array with size n by p .
<code>z_dims</code>	is the latent dimensions for z_0, z_1, z_2, z_3 respectively. Total dimension should be much smaller than the dimension of covariates v . Default: <code>c(3,3,6,6)</code>
<code>output_dir</code>	is the folder to save the results including model hyperparameters and the estimated causal effect. Default is <code>"."</code> .
<code>dataset</code>	is the name for the input data. Default: <code>"myData"</code> .
<code>lr</code>	is the learning rate. Default: <code>0.0002</code> .
<code>bs</code>	is the batch size. Default: <code>32</code> .
<code>alpha</code>	is the coefficient for the reconstruction loss. Default: <code>1</code> .
<code>beta</code>	is the coefficient for the MSE loss of x and y . Default: <code>1</code> .
<code>gamma</code>	is the coefficient for the gradient penalty loss. Default: <code>10</code> .
<code>g_d_freq</code>	is the iteration frequency between training generator and discriminator in the Roundtrip framework. Default: <code>5</code> .
<code>g_units</code>	is the list of hidden nodes in the generator/decoder network. Default: <code>c(64,64,64,64,64)</code> .
<code>e_units</code>	is the list of hidden nodes in the encoder network. Default: <code>c(64,64,64,64,64)</code> .
<code>f_units</code>	is the list of hidden nodes in the f network for predicting y . Default: <code>c(64,32,8)</code> .
<code>h_units</code>	is the list of hidden nodes in the h network for predicting x . Default: <code>c(64,32,8)</code> .
<code>dv_units</code>	is the list of hidden nodes in the discriminator for distribution match v . Default: <code>c(64,32,8)</code> .
<code>dz_units</code>	is the list of hidden nodes in the discriminator for distribution match z . Default: <code>c(64,32,8)</code> .
<code>save_model</code>	whether to save the trained model. Default: <code>FALSE</code> .
<code>save_res</code>	whether to save the results during training. Default: <code>FALSE</code> .
<code>binary_treatment</code>	whether the treatment is binary or continuous. Default: <code>TRUE</code> .
<code>use_z_rec</code>	whether to use the reconstruction loss for z . Default: <code>TRUE</code> .
<code>use_v_gan</code>	whether to use the GAN training for v . Default: <code>TRUE</code> .
<code>random_seed</code>	is the random seed to fix randomness. Default: <code>123</code> .
<code>n_iter</code>	is the training iterations. Default: <code>30000</code> .
<code>normalize</code>	whether apply normalization to covariates. Default: <code>FALSE</code> .
<code>x_min</code>	ADRF start value. Default: <code>NULL</code>
<code>x_max</code>	ADRF end value. Default: <code>NULL</code>

Value

`causalegm` returns an object of `class` `"causalegm"`.

An object of class `"causalegm"` is a list containing the following:

<code>causal_pre</code>	the predicted causal effects, which are individual causal effects (ITEs) in binary treatment settings and dose-response values in continuous treatment settings.
<code>getCATE</code>	the method for getting the conditional average treatment effect (CATE). It takes covariates v as input.
<code>predict</code>	the method for outcome function. It takes treatment x and covariates v as inputs.

References

Qiao Liu, Zhongren Chen, Wing Hung Wong. CausalEGM: a general causal inference framework by encoding generative modeling. *arXiv preprint arXiv:2212.05925*, 2022.

Examples

```
#Generate a simple simulation data.
n <- 1000
p <- 10
v <- matrix(rnorm(n * p), n, p)
x <- rbinom(n, 1, 0.4 + 0.2 * (v[, 1] > 0))
y <- pmax(v[, 1], 0) * x + v[, 2] + pmin(v[, 3], 0) + rnorm(n)
model <- causalegm(x=x, y=y, v=v, n_iter=3000)
paste("The average treatment effect (ATE):", round(model$ATE, 2))
```

get_est

Make predictions with causalEGM model.

Description

When `x` is `NULL`, the conditional average treatment effect (CATE), namely $\tau(v)$, is estimated using a trained causalEGM model. When `x` is provided, estimating the potential outcome `y` given treatment `x` and covariates `v` using a trained causalEGM model.

Usage

```
get_est(object, v, x = NULL)
```

Arguments

<code>object</code>	An object of class "causalegm".
<code>v</code>	is the covariates, two-dimensional array with size <code>n</code> by <code>p</code> .
<code>x</code>	is the optional treatment variable, one-dimensional array with size <code>n</code> . Defaults to <code>NULL</code> .

Value

Vector of predictions.

Examples

```
#Generate a simple simulation data.
n <- 1000
p <- 10
v <- matrix(rnorm(n * p), n, p)
x <- rbinom(n, 1, 0.4 + 0.2 * (v[, 1] > 0))
y <- pmax(v[, 1], 0) * x + v[, 2] + pmin(v[, 3], 0) + rnorm(n)
model <- causalegm(x=x, y=y, v=v, n_iter=3000)
n_test <- 100
v_test <- matrix(rnorm(n_test * p), n_test, p)
x_test <- rbinom(n_test, 1, 0.4 + 0.2 * (v_test[, 1] > 0))
pred_cate <- get_est(model, v = v_test) # CATE estimate
pred_y <- get_est(model, v = v_test, x = x_test) # y given treatment x plus covariates v
```

install_causalegm *Install the python CausalEGM package*

Description

Install the python CausalEGM package

Usage

```
install_causalegm(method = "auto", pip = TRUE)
```

Arguments

method	default "auto"
pip	boolean flag, default TRUE

Value

No return value

Index

causalegm, 2

class, 3

get_est, 4

install_causalegm, 5