Package ‘RcausalEGM’

March 28, 2023

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

R topics documented:

causalegm *Main function for estimating causal effect in either binary or continuous treatment settings.*

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.

v is the covariates, two-dimensional array with size n by p.

z_dims 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: c(3,3,6,6)

output_dir is the folder to save the results including model hyperparameters and the estimated causal effect. Default is ".".

dataset is the name for the input data. Default: "myData".

lr is the learning rate. Default: 0.0002.

bs is the batch size. Default: 32.

alpha is the coefficient for the reconstruction loss. Default: 1.

beta is the coefficient for the MSE loss of x and y. Default: 1.

gamma is the coefficient for the gradient penalty loss. Default: 10.

g_d_freq is the iteration frequency between training generator and discriminator in the Roundtrip framework. Default: 5.

g_units is the list of hidden nodes in the generator/decoder network. Default: c(64,64,64,64,64).

e_units is the list of hidden nodes in the encoder network. Default: c(64,64,64,64,64).

f_units is the list of hidden nodes in the f network for predicting y. Default: c(64,32,8).

h_units is the list of hidden nodes in the h network for predicting x. Default: c(64,32,8).

dv_units is the list of hidden nodes in the discriminator for distribution match v. Default: c(64,32,8).

dz_units is the list of hidden nodes in the discriminator for distribution match z. Default: c(64,32,8).

save_model whether to save the trained model. Default: FALSE.

save_res whether to save the results during training. Default: FALSE.

binary_treatment
whether the treatment is binary or continuous. Default: TRUE.

use_z_rec whether to use the reconstruction loss for z. Default: TRUE.

use_v_gan whether to use the GAN training for v. Default: TRUE.

random_seed is the random seed to fix randomness. Default: 123.

n_iter is the training iterations. Default: 30000.

normalize whether apply normalization to covariates. Default: FALSE.

x_min ADRF start value. Default: NULL

x_max ADRF end value. Default: NULL

Value

causalegm returns an object of class "causalegm".

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

causal_pre the predicted causal effects, which are individual causal effects (ITEs) in binary treatment settings and dose-response values in continous treatment settings.

getCATE the method for getting the conditional average treatment effect (CATE).It takes covariates v as input.

predict 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

`object` An object of class `"causalegm"`.

`v` is the covariates, two-dimensional array with size n by p.

`x` is the optional treatment variable, one-dimensional array with size n. Defaults to NULL.

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