Package 'LVGP'

January 20, 2025

Type Package

Title Latent Variable Gaussian Process Modeling with Qualitative and Quantitative Input Variables

Version 2.1.5

Author Siyu Tao, Yichi Zhang, Daniel W. Apley, Wei Chen

Maintainer Siyu Tao <siyutao2020@u.northwestern.edu>

Description

Fit response surfaces for datasets with latent-variable Gaussian process modeling, predict responses for new inputs, and plot latent variables locations in the latent space (only 1D or 2D).
The input variables of the datasets can be quantitative, qualitative/categorical or mixed.
The output variable of the datasets is a scalar (quantitative).
The optimization of the likelihood function is done using a successive approximation/relaxation algorithm similar to another GP modeling package ``GPM".
The modeling method is published in ``A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors"
by Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (2018) <arXiv:1806.07504>. The package is developed in IDEAL of Northwestern University.

License GPL-2

Encoding UTF-8

LazyData true

Imports lhs(>= 0.14), randtoolbox(>= 1.17)

Depends R (>= 3.4.0), stats (>= 3.2.5), parallel (>= 3.2.5)

Repository CRAN

RoxygenNote 6.1.1

NeedsCompilation no

Date/Publication 2019-01-11 07:50:03 UTC

Contents

`corr_mat` *The Function for Constructing the Correlation Matrix in* `LVGP` *Package*

Description

Builds the correlation matrix given two datasets, and the type and parameters of the correlation function.

Usage

```
corr_mat(X1, X2, phi_full)
```

Arguments

`X1, X2`	Matrices containing the data points. The rows and columns of both `X1` and `X2` denote individual observation settings and dimension, respectively.
`phi_full`	The vector storing all the scale (aka roughness) parameters of the correlation function. See `reference 1`.

Value

R The Correlation matrix with size `nrow(X1)`-by-`nrow(X2)`. See here.

Note

This function is **NOT** exported once the LVGP package is loaded.

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

See Also

`LVGP_fit` to see how a GP model can be fitted to a training dataset.
`LVGP_predict` to use the fitted LVGP model for prediction.
`LVGP_plot` to plot the features of the fitted model.

Examples

```
# see the examples in the documentation of the function LVGP_fit.
```

LVGP_fit *The Fitting Function of* LVGP *Package*

Description

Fits a latent-variable Gaussian process (LVGP) model to a dataset as described in reference 1. The input variables can be quantitative or qualitative/categorical or mixed. The output variable is quantitative and scalar.

Usage

```
LVGP_fit(X, Y, ind_qual = NULL, dim_z = 2, eps = 10^(seq(-1, -8,
  length.out = 15)), lb_phi_ini = -2, ub_phi_ini = 2,
  lb_phi_lat = -8, ub_phi_lat = 3, lb_z = -3, ub_z = 3,
  n_opt = 8, max_iter_ini = 100, max_iter_lat = 20, seed = 123,
  progress = FALSE, parallel = FALSE, noise = FALSE)
```

Arguments

X	Matrix or data frame containing the inputs of training data points. Each row is a data point.
Y	Vector containing the outputs of training data points
ind_qual	Vector containing the indices of columns of qualitative/categorical variables
dim_z	Dimensionality of latent space, usually 1 or 2 but can be higher
eps	The vector of smallest eigen values that the correlation matrix is allowed to have, which determines the nugget added to the correlation matrix.
lb_phi_ini, ub_phi_ini	The initial lower and upper search bounds of the scale/roughness parameters (phi) of quantitative variables
lb_phi_lat, ub_phi_lat	The later lower and upper search bounds of the scale/roughness parameters (phi) of quantitative variables
lb_z, ub_z	The lower and upper search bounds of the latent parameters (z) of qualitative variables
n_opt	The number of times the log-likelihood function is optimized
max_iter_ini	The maximum number of iterations for each optimization run for largest (first) eps case
max_iter_lat	The maximum number of iterations for each optimization run for after first eps cases
seed	An integer for the random number generator. Use this to make the results reproducible.
progress	The switch determining whether to print function run details
parallel	The switch determining whether to use parallel computing
noise	The switch for whether the data are assumed noisy

Value

A model of class "LVGP model" list of the following items:

- `quant_param` A list containing the estimated parameter `phi` and its search bounds for quantitative variables
- `qual_param` A list containing the estimated parameter `z` and its dimensionality, vectorized form and search bounds for qualitative variables
- `data` A list containing the fitted dataset in verbose format
- `fit_detail` A list of more detailed variables for fitting and prediction process
- `optim_hist` Optimization history
- `setting` Settings for the optimization and fitting process

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

See Also

`optim` for the details on `L-BFGS-B` algorithm used in optimization.
`LVGP_predict` to use the fitted LVGP model for prediction.
`LVGP_plot` to plot the features of the fitted model.

Examples

```
# Math example with 2 quantitative and 1 qualitative variables (dataset included in the package):
#     Fit a model (with default settings) and evaluate the performance
#     by computing the root mean squared error (RMSE) in prediction.
#     Also, plot the latent variable parameters.
X_tr <- math_example$X_tr
Y_tr <- math_example$Y_tr
X_te <- math_example$X_te
Y_te <- math_example$Y_te
n_te <- nrow(X_te)
model <- LVGP_fit(X_tr, Y_tr, ind_qual = c(3))
output <- LVGP_predict(X_te, model)
Y_hat <- output$Y_hat
RRMSE <- sqrt(sum((Y_hat-Y_te)^2)/n_te)/(max(Y_te)-min(Y_te))
LVGP_plot(model)
```

LVGP_plot *The Plotting Function of* LVGP *Package*

Description

Plots the qualitative/categorical variable levels in the latent space (only for 1D or 2D cases). If the qualitative/categorical variables are not specified, all the qualified variables will be plotted. See Arguments for more details on the options.

Usage

```
LVGP_plot(model, ind_qual_plot = NULL)
```

Arguments

model	The LVGP model fitted by LVGP_fit
ind_qual_plot	An array of index (indices) of the qualitative/categorical variable(s) to be plotted. Default is NULL, in which case all the qualitative/categorical variables will be plotted.

Note

This plot function **only** works for 1D or 2D latent spaces.

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

See Also

LVGP_fit to fit LVGP model to the datasets.
LVGP_predict to use the fitted LVGP model for prediction.

Examples

```
# see the examples in the documentation of the function LVGP_fit.
```

LVGP_predict *The Prediction Function of* LVGP *Package*

Description

Predicts the output and associated uncertainties of the GP model fitted by LVGP_fit.

Usage

```
LVGP_predict(X_new, model, MSE_on = 0)
```

Arguments

X_new	Matrix or vector containing the input(s) where the predictions are to be made. Each row is an input vector.
model	The LVGP model fitted by LVGP_fit.
MSE_on	A scalar indicating whether the uncertainty (i.e., mean squared error MSE) is calculated. Set to a non-zero value to calculate MSE.

Value

A prediction list containing the following components:

- Y_hat A vector containing the mean prediction values
- MSE A vector containing the prediction uncertainty (i.e., the covariance or covariance matrix for the output(s) at prediction location(s))

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

See Also

LVGP_fit to fit LVGP model to the datasets.
LVGP_plot to plot the features of the fitted model.

Examples

```
# see the examples in the documentation of the function LVGP_fit.
```

`math_example` *Dataset for the example in function 'LVGP_fit'*

Description

Data are sampled from the modified math function based on the first example in the paper listed in codereferences. There are still 2 quantitative and 1 qualitative variables, but the qualitative variable has only 3 levels. For each level, there are 8 training data points and 30 testing data points, all generated with Latin hypercube sampling. In total, there are 24 training data points and 90 testing data points.

Usage

```
data(math_example)
```

Format

A named list containing training and test data:

"X_tr" 24-by-3 matrix for 24 training data inputs, 3rd column being the qualitative variable

"Y_tr" 24-by-1 matrix for 24 training data outputs

"X_te" 90-by-3 matrix for 90 testing data inputs, 3rd column being the qualitative variable

"Y_te" 90-by-1 matrix for 90 testing data outputs

Source

The dataset can be generated with the code at the end of this description file.

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

Examples

```
data(math_example)
X_tr <- math_example$X_tr
Y_tr <- math_example$Y_tr
X_te <- math_example$X_te
Y_te <- math_example$Y_te
```

neg_log_l *The Negative Log-Likelihood Function in* LVGP *Package*

Description

Calculates the negative log-likelihood (excluding all the constant terms) as described in reference 1.

Usage

```
neg_log_l(hyperparam, p_quant, p_qual, lvs_qual, n_lvs_qual, dim_z,
  X_quant, X_qual, Y, min_eig, k, M)
```

Arguments

hyperparam	Hyperparameters of the LVGP model
p_quant	Number of quantative variables
p_qual	Number of qualitative variables
lvs_qual	Levels of each qualitative variable
n_lvs_qual	Number of levels of each qualitative variable
dim_z	Dimensionality of latent variables, usually 1 or 2
X_quant	Input data of quantative variables
X_qual	Input data of qualitative variables
Y	Vector containing the outputs of data points
min_eig	The smallest eigen value that the correlation matrix is allowed to have, which determines the nugget added to the correlation matrix.
k	Number of data points, nrow(X_quant) or nrow(X_qual)
M	Vector of ones with length k

Details

LVGP_fit calls this function as its optimization objective function.

Value

The negative log-likelihood (excluding all the constant terms) value.

Note

This function is **NOT** exported once the package is loaded.

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

See Also

LVGP_fit to see how a GP model can be fitted to a training dataset.
LVGP_predict to use the fitted LVGP model for prediction.
LVGP_plot to plot the features of the fitted model.

Examples

```
# see the examples in the documentation of the function LVGP_fit.
```

---

to_latent *The Function for Transforming Qualitative/Categorical Variables into Latent Variables in* LVGP *Package*

Description

Transforms qualitative/categorical variables into latent variables.

Usage

```
to_latent(X_qual, lvs_qual, n_lvs_qual, p_qual, z_vec, dim_z, k)
```

Arguments

X_qual	Matrix or data frame containing (only) the qualitative/categorical data.
lvs_qual	List containing levels of each qualitative variable
n_lvs_qual	Number of levels of each qualitative variable
p_qual	Number of qualitative variables
z_vec	Latent variable parameters, i.e., latent variable values for each level of qualitative/categorical variables
dim_z	Dimensionality of latent variables, usually 1 or 2
k	Number of data points, equal to nrow(X_qual)

Value

Matrix containing transformed data

Note

This function is **NOT** exported once the LVGP package is loaded.

References

1. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors", Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W. Apley (arXiv)

See Also

LVGP_fit to see how a GP model can be fitted to a training dataset.
LVGP_predict to use the fitted LVGP model for prediction.
LVGP_plot to plot the features of the fitted model.

Examples

```
# see the examples in the documentation of the function LVGP_fit.
```

Index