Documentation for the R code to implement the FAME methodology in “Functional Adaptive Model Estimation”

The R script contains the function to fit the FAME estimator. The main function is `fame`.

**Description**

This function fits the FAME estimator for regression with a functional covariate and a scalar response according to the procedure outlined in the paper. It also provides an option to include scalar covariates.

**fame**

This is the main function that fits the FAME estimator.

**Arguments**

`data` : a list with four or five components defined as follows

- `x` : a vector \((x_1^T, x_2^T, ..., x_N^T)^T\) that contains each of the evaluated points of the functional predictor. Each \(x_i\) is a vector containing the observations from curve \(i\). Note that each curve can have observations at different time points.
- `y` : a length \(N\) vector that contains the scalar response
- `z` : a matrix that contains the scalar predictors (optional)
- `time` : a vector that specifies the time points that the corresponding component of `x` is measured at. It has the same length as `x`.
- `curve` : a vector of the form \((1, 1, ..., 1, 2, 2, ...2...)\), assuming that the first points in `x` come from curve 1, then the next points from curve 2 etc. It has the same length as `x`.
- `q` : dimension of cubic B-spline basis \((q \geq 4)\)
- `p` : a lower dimension that the basis coefficients can be projected onto (defaults to 4)
- `r` : number of additive components (integer, \(r \geq 1\))
- `grid` : observation grid of the functional predictor. It includes all `time` elements in `data` and also other time points that future data may be observed at.
- `t_range` : a vector of length 2 that specifies the lower and upper bound of the time domain of the functional predictor
- `maxit` : the maximum number of training iterations (defaults to 3)
- `family` : family of link functions (defaults to `gaussian()`)

1
**pertfit** : an optional initial fit provided by the user (defaults to NULL)

- **pc** : a logical value indicating if FAME is fitted using FPCs or not (defaults to FALSE)
- **run** : number of iterations to compute the initial fit (defaults to 4)
- **tol** : convergence criterion to stop training on small change in the deviance of log likelihood (defaults to .001)

**Value**

A list with the following components

- **BaseS** : a matrix of cubic B-spline basis evaluations. Its dimension is the length of grid by q.
- **W** : a q by p matrix that projects the basis to a lower dimensional space
- **alpha** : a N by p matrix of coefficients that project x to a lower dimensional space
- **Theta** : a p by r matrix of coefficients that maps lower dimensional projections of x to $\int x(t)\beta_k(t)dt$ for $k = 1, \ldots, r$. Each column corresponds to the coefficients for each additive component.

**gamfit** : an object returned by `gam`. This is the fit regressing y on r projections of x.

**fame.predict**

This is the function to make predictions given test data and an object returned by `fame`.

**Arguments**

- **obj** : an object returned by `fame`
- **newdata** : a list for test data. It needs to be specified in the same format as `data` in `fame`.
- **type** : type of returned predictions, chosen from "response", "terms", "link", "lpmatrix", and "iterms" (defaults to "response", see documentation of `predict.gam`)

**Value**

This function returns predicted values depending on **type**. When **type** = "response", it returns a vector of predicted values of the response.