

## GAM 一般化加法モデル ( Generalized Additive Model )

- ・分布が非線形な場合、非線形な分布を表す一つの曲線を求めようというのではなく、
- ・分割して、より単純な分布を足したものとしてとらえる。

mgcv パッケージ : Mixed GAM Computation Vehicle with Automatic Smoothness Estimation

<https://cran.r-project.org/web/packages/mgcv/index.html>

説明変数を s() に入れる ( s は smooth 平滑化 )

- ・ オプションで k の値を明示的に設定
  - ・ knot (全体を区分に分割する数)
    - ・ 数が多くて細かすぎると、線がぐにゃぐにゃになる
  - ・ デフォルトは自動で変数の数に応じて設定される
- ・ オプションで bs で曲線の描き方を設定
  - ・ tp (thin plate regression spline)
  - ・ cr (cubic regression spline)
- ・ オプション sp で平滑化パラメタの数を設定
  - ・ 少ないほうが好ましいが、少ないと線がぐにゃぐにゃになる

smooth terms の重要性

- ・ edf: Effective Degrees of Freedom (有効自由度)
  - ・ データに適合するのに使った自由度の数
  - ・ 高いほどデータに適合している
- ・ Ref.df: Reference Degrees of Freedom (参照自由度)
  - ・ 基準となる自由度
  - ・ 複雑さを示す
- ・ F 統計量
  - ・ 大きいほどスムーズ項がモデルに寄与することを示す

モデルの評価指標

- ・ R-sq.(adj) (調整済み決定係数)
  - ・ Deviance explained (説明された偏差)
- ・ GCV (Generalized Cross Validation) : 予測精度
- ・ Scale est. (スケールの推定値)

plot.gam {mgcv}

- ・ オプション
  - ・ residuals=F で残差表示
  - ・ se=F で標準誤差を非表示
  - ・ pages=1 ですべてを 1 ページに
  - ・ jit=T でジッター表示
  - ・ shade=T で信頼区間を影表示
    - ・ shade.col="色"
- ・ y 軸は実測値ではなく、予測値のベースラインが 0 となる
- ・ 実測値を確認するには、

- ・ できたモデルに、データを入れて、「予測値」をプロットする
- ・ `plot.xy()`
  - ・ `type` でプロットのタイプを指定
    - ・ `p`: points
    - ・ `l`: lines
    - ・ `b`: both
    - ・ `o`: overplotted points and lines
    - ・ `s`: stair steps
    - ・ `h`: histogram
  - ・ ただし、データの並び順が適切にならずに、線で結ぶと、順につながらずに、行ったり来たりする結び方をする（原因不明）
    - ・ `x` と `y` でデータフレームを作った上で、`x` でソートしなおせばきれいに線が引ける

## gam パッケージ

<https://cran.r-project.org/web/packages/gam/index.html>

## References

<https://qiita.com/tabintone/items/96afd222d92e876c51d1>

[https://qiita.com/purple\\_jp/items/5d138f5652f469dbfe9f](https://qiita.com/purple_jp/items/5d138f5652f469dbfe9f)

[https://www.jstage.jst.go.jp/article/jbhmk/34/1/34\\_1\\_111/pdf](https://www.jstage.jst.go.jp/article/jbhmk/34/1/34_1_111/pdf)