Additive Model and Generalized Additive Model

Luwei Ying

Washington University in St. Louis

December 7, 2017

Outline

Why do we need additive models and generalized additive models?

Additive model (AM)

Generalized Additive model (GAM)

Example of GAM: Supreme Court Overrides

Motivation

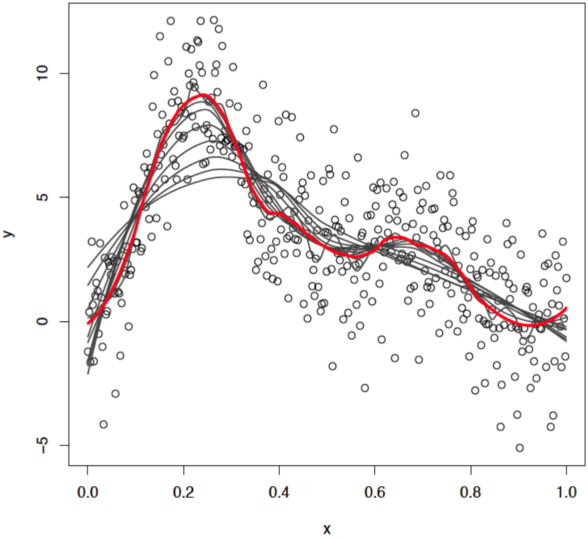

Figure 1: Non-parametric Regressions

Additive model

One additional assumption: the effects of predictor variables (X_1, X_2, ...) are additive.

Additive model

One additional assumption: the effects of predictor variables (X_1, X_2, ...) are additive.
Parametric linear regression model

$$Y_i = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

Additive model

One additional assumption: the effects of predictor variables (X_1, X_2, ...) are additive.
Parametric linear regression model

$$Y_i = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

Non-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Additive model

One additional assumption: the effects of predictor variables (X_1, X_2, ...) are additive.
Parametric linear regression model

$$Y_i = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

Non-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Semi-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_j(X_j) + \beta_1 X_{j+1} + \cdots + \beta_k X_k + \varepsilon$$

Additive model

One additional assumption: the effects of predictor variables (X_1, X_2, ...) are additive.
Parametric linear regression model

$$Y_i = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

Non-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Semi-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_j(X_j) + \beta_1 X_{j+1} + \cdots + \beta_k X_k + \varepsilon$$

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

1. Set $\alpha = \bar{Y}$ and $S_j = X$ as starting values for $j = 1, ..., k$.

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

1. Set $\alpha = \bar{Y}$ and $S_j = X$ as starting values for $j = 1, ..., k$.
2. Calculate the partial residual for each X:
   $\hat{e}_p^j = Y_i - \sum_{i=2}^{k} S_j - \alpha$.

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

1. Set $\alpha = \bar{Y}$ and $S_j = X$ as starting values for $j = 1, ..., k$.
2. Calculate the partial residual for each X: $\hat{e}_p^j = Y_i - \sum_{i=2}^{k} S_j - \alpha$.
3. Smooth $\hat{e}_p^j$ on X_1.

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

1. Set $\alpha = \bar{Y}$ and $S_j = X$ as starting values for $j = 1, ..., k$.
2. Calculate the partial residual for each X: $\hat{e}_p^j = Y_i - \sum_{i=2}^{k} S_j - \alpha$.
3. Smooth $\hat{e}_p^j$ on X_1.
4. Replace the covariate X_1 in S_j with the predictions from the smoothed fit at the values of X_i.

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

1. Set $\alpha = \bar{Y}$ and $S_j = X$ as starting values for $j = 1, ..., k$.
2. Calculate the partial residual for each X:
   $\hat{e}_p^j = Y_i - \sum_{i=2}^{k} S_j - \alpha$.
3. Smooth $\hat{e}_p^j$ on X_1.
4. Replace the covariate X_1 in S_j with the predictions from the smoothed fit at the values of X_i.
5. Repeat steps $2 - 4$ for each X from 2 to k.

Estimation: Backfitting

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Let S_j denote a matrix where each column represents each estimate of f_k, and X be a model matrix where each column is one of the X covariates.

1. Set $\alpha = \bar{Y}$ and $S_j = X$ as starting values for $j = 1, ..., k$.
2. Calculate the partial residual for each X: $\hat{e}_p^j = Y_i - \sum_{i=2}^{k} S_j - \alpha$.
3. Smooth $\hat{e}_p^j$ on X_1.
4. Replace the covariate X_1 in S_j with the predictions from the smoothed fit at the values of X_i.
5. Repeat steps $2 - 4$ for each X from 2 to k.
6. Calculate the model residual sum of squares as:

$$RSS = \sum_{i=1}^{n} \left[\left(Y_i - \sum_{i=1}^{k} S_j \right)^2 \right]$$

Additive model

One additional assumption: the effects of predictor variables (X_1, X_2, ...) are additive.
Parametric linear regression model

$$Y_i = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

Non-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_k(X_k) + \varepsilon$$

Semi-parametric regression model with additive effects

$$Y_i = \alpha + f_1(X_1) + \cdots + f_j(X_j) + \beta_1 X_{j+1} + \cdots + \beta_k X_k + \varepsilon$$

Generalized Additive model (GAM)

GLM:

$$g(\eta_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{2i}$$

Generalized Additive model (GAM)

GLM:

$$g(\eta_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{2i}$$

GAM:

$$g(\eta_i) = \alpha + f_1(X_{i1}) + f_2(X_{2i})$$

Generalized Additive model (GAM)

GLM:

$$g(\eta_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{2i}$$

GAM:

$$g(\eta_i) = \alpha + f_1(X_{i1}) + f_2(X_{2i})$$

We can also use the backfitting algorithm to estimate the smoothing components in GAMs.

Count Models: Supreme Court Overrides

Count Models: Supreme Court Overrides

- Y: the number of Congressional acts overturned by the Supreme Court each year (1789 – 1990)

Count Models: Supreme Court Overrides

- Y: the number of Congressional acts overturned by the Supreme Court each year (1789 – 1990)
- X1: Justice Tenure – the average number of years served by the justices sitting on the Court that year

Count Models: Supreme Court Overrides

- Y: the number of Congressional acts overturned by the Supreme Court each year (1789 – 1990)
- X1: Justice Tenure – the average number of years served by the justices sitting on the Court that year
- X2: Unified Congress – a dummy variable that indicates if both chambers of Congress are controlled by the same party

Count Models: Supreme Court Overrides

- Y: the number of Congressional acts overturned by the Supreme Court each year (1789 – 1990)
- X1: Justice Tenure – the average number of years served by the justices sitting on the Court that year
- X2: Unified Congress – a dummy variable that indicates if both chambers of Congress are controlled by the same party
- X3: Congress Counter – which year (to control for the institutionalization of judicial review and to capture the general propensity of the Court to overturn Congress)

Count Models: Supreme Court Overrides

- Y: the number of Congressional acts overturned by the Supreme Court each year (1789 – 1990)
- X1: Justice Tenure – the average number of years served by the justices sitting on the Court that year
- X2: Unified Congress – a dummy variable that indicates if both chambers of Congress are controlled by the same party
- X3: Congress Counter – which year (to control for the institutionalization of judicial review and to capture the general propensity of the Court to overturn Congress) linear? quadratic?

Count Models: Supreme Court Overrides

- Y: the number of Congressional acts overturned by the Supreme Court each year (1789 – 1990)
- X1: Justice Tenure – the average number of years served by the justices sitting on the Court that year
- X2: Unified Congress – a dummy variable that indicates if both chambers of Congress are controlled by the same party
- X3: Congress Counter – which year (to control for the institutionalization of judicial review and to capture the general propensity of the Court to overturn Congress)
  linear? quadratic?
  up & down & up & down & up & down & up & down?

Supreme Court Overrides: R code for the models

```
library('mgcv')

#Parametric Model
mod.1 <- gam(nulls ~ tenure + congress + unified,
             data=Scourt, family=poisson)
#GAM
mod.2 <- gam(nulls ~ tenure + s(congress, bs="cr") + unified,
             data=Scourt, family=poisson)
```

Supreme Court Overrides: Result

Table 6.1 A comparison of parametric and semiparametric models of Supreme Court overrides.

	Supreme Court overrides (parametric)	Supreme Court overrides (semiparametric)
Justice Tenure	0.07^{*}	0.19^{***}
	(0.03)	(0.05)
Unified Congress	0.14	0.18
	(0.24)	(0.28)
Congress Counter	0.03^{***}	$-^{***}$
	(0.003)	
Constant	-2.7^{***}	-3.31^{***}
	(0.52)	(0.77)
Deviance explained	44%	67%
LR Test p-value		0.00

Likelihood ratio test against previous model in the table.

Standard errors in parentheses. Two-tailed tests.

$^{*}p$-value < 0.05 ** p-value < 0.01 *** p-value < 0.001

Supreme Court Overrides: Non-parametric Estimate

```
plot(mod.2, rug=FALSE, ylab="Propensity to Overturn
    Congressional Acts", xlab="Congress", shift=-3.31)
```

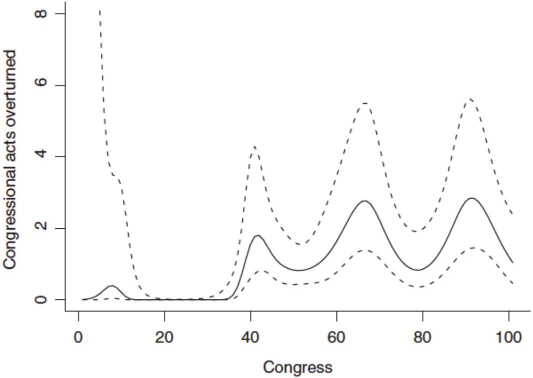

Figure 6.7 Nonparametric estimate for Congressional counter variable.

Supreme Court Overrides: How other term is influenced

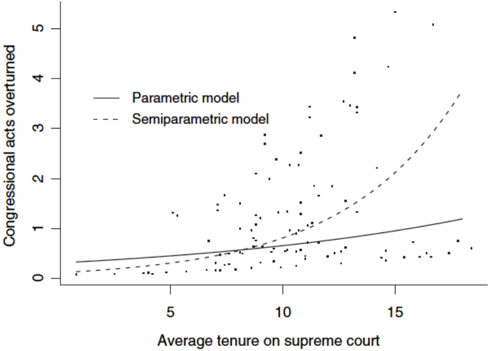

Figure 6.6 The difference in the effect of tenure across parametric and semi-parametric models.

Conclusion

- Additive model and generalized additive model

Conclusion

- Additive model and generalized additive model
- Non-parametric $\rightarrow$ Semi-parametric

Conclusion

- Additive model and generalized additive model
- Non-parametric $\rightarrow$ Semi-parametric
- Reconsider the predictors in your research: is there nonlinearity?

Conclusion

- Additive model and generalized additive model
- Non-parametric $\rightarrow$ Semi-parametric
- Reconsider the predictors in your research: is there nonlinearity?
- If YES, you SHOULD use these models

Conclusion

- Additive model and generalized additive model
- Non-parametric $\rightarrow$ Semi-parametric
- Reconsider the predictors in your research: is there nonlinearity?
- If YES, you SHOULD use these models
- Because undiagnosed nonlinearity can infect other terms in the model