MONASH University

ETC3250

Business Analytics

3. Regression

10 August 2015

Outline

1. **Revision: multiple regression**
2. Matrix formulation
3. Splines

Revision: Multiple regression

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

- Each $X_{j,i}$ is numerical and is called a "predictor".
- The coefficients $\beta_1, \ldots, \beta_p$ measure the effect of each predictor after taking account of the effect of all other predictors in the model.
- Predictors may be transforms of other predictors. e.g., $X_2 = X_1^2$.
- The model describes a line, plane or hyperplane in the predictors.

Revision: Multiple regression

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

- Each $X_{j,i}$ is numerical and is called a "predictor".
- The coefficients $\beta_1, \dots, \beta_p$ measure the effect of each predictor after taking account of the effect of all other predictors in the model.
- Predictors may be transforms of other predictors. e.g., $X_2 = X_1^2$.
- The model describes a line, plane or hyperplane in the predictors.

Revision: Multiple regression

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

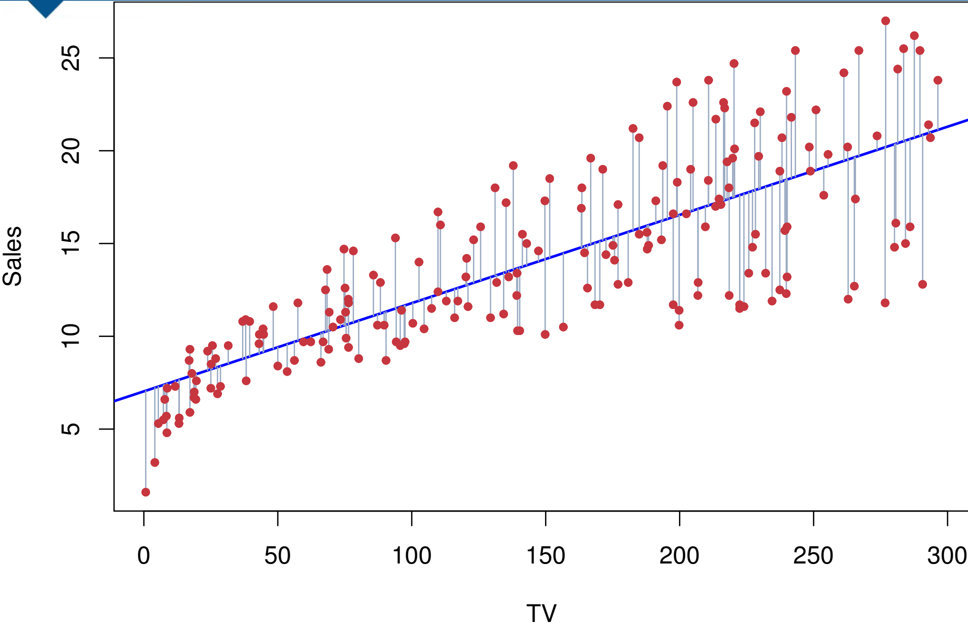
- Each $X_{j,i}$ is numerical and is called a “predictor”.
- The coefficients $\beta_1, \ldots, \beta_p$ measure the effect of each predictor after taking account of the effect of all other predictors in the model.
- Predictors may be transforms of other predictors. e.g., $X_2 = X_1^2$.
- The model describes a line, plane or hyperplane in the predictors.

Revision: Multiple regression

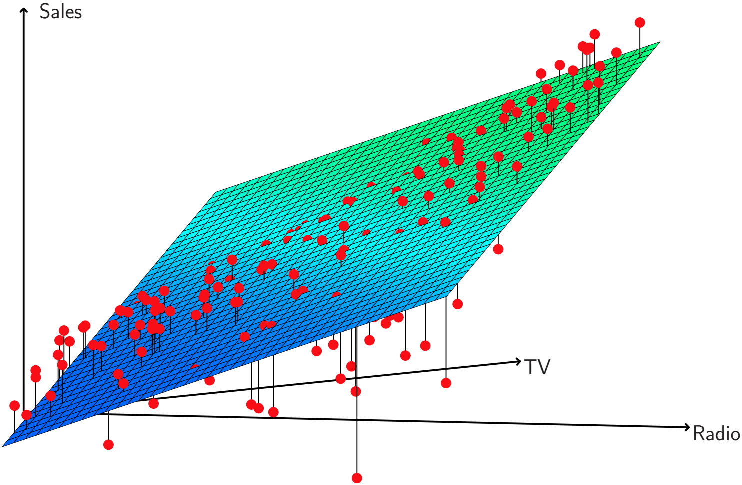
$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

- Each $X_{j,i}$ is numerical and is called a "predictor".
- The coefficients $\beta_1, \dots, \beta_p$ measure the effect of each predictor after taking account of the effect of all other predictors in the model.
- Predictors may be transforms of other predictors. e.g., $X_2 = X_1^2$.
- The model describes a line, plane or hyperplane in the predictors.

Revision: Multiple regression



Revision: Multiple regression



Revision: Multiple regression

- Dummy variables
- OLS
- R^2
- se of coefficients
- residual standard error
- F statistic

Revision: Multiple regression

- Dummy variables
- OLS
- R^2
- se of coefficients
- residual standard error
- F statistic

Revision: Multiple regression

- Dummy variables
- OLS
- R^2
- se of coefficients
- residual standard error
- F statistic

Revision: Multiple regression

- Dummy variables
- OLS
- R^2
- se of coefficients
- residual standard error
- F statistic

Revision: Multiple regression

- Dummy variables
- OLS
- R^2
- se of coefficients
- residual standard error
- F statistic

Revision: Multiple regression

- Dummy variables
- OLS
- R^2
- se of coefficients
- residual standard error
- F statistic

Important questions

1. Is at least one of the predictors useful in predicting the response?
2. Do all the predictors help to explain Y, or is only a subset of the predictors useful?
3. How well does the model fit the data?
4. Given a set of predictor values, what response value should we predict and how accurate is our prediction?

Credit scores

Banks score loan customers based on a lot of personal information. A sample of 500 customers from an Australian bank provided the following information.

Score	Savings $'000	Income $'000	Time current address Months	Time current job Months
39.40	0.01	111.17	27	8
51.79	0.65	56.40	29	33
32.82	0.75	36.74	2	16
57.31	0.62	55.99	14	7
37.17	4.13	62.04	2	14
33.69	0.00	43.75	7	7
25.56	0.94	79.01	4	11
32.04	0.00	45.41	3	3
41.34	4.26	55.22	16	18
⋮	⋮	⋮	⋮	⋮

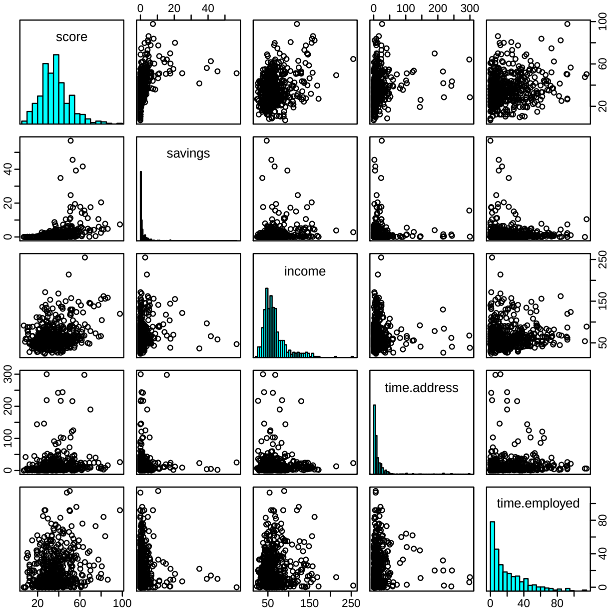
Credit scores

Banks score loan customers based on a lot of personal information. A sample of 500 customers from an Australian bank provided the following information.

Score	Savings $’000	Income $’000	Time current address Months	Time current job Months
39.40	0.01	111.17	27	8
51.79	0.65	56.40	29	33
32.82	0.7	[illegible]	[illegible]	6
57.31	0.6	[illegible]	[illegible]	7
37.17	4.1	[illegible]	[illegible]	4
33.69	0.0	[illegible]	[illegible]	7
25.56	0.94	79.01	4	11
32.04	0.00	45.41	3	3
41.34	4.26	55.22	16	18
⋮	⋮	⋮	⋮	⋮

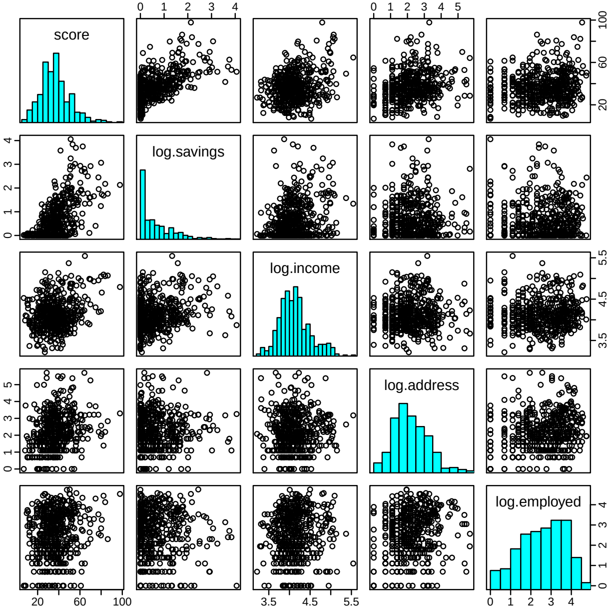
Can we use only savings, income, time @ address and time employed to predict the score of a new customer?

Credit scores



Credit scores

- Taking logarithms reduces the skewness in the predictor variables.
- Because of zeros, I used $\log(x+1)$.



Credit scores

Proposed model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$$

$$\begin{aligned} \text{where } \; Y &= \text{Credit score}, \\ X_1 &= \text{log savings}, \\ X_2 &= \text{log income}, \\ X_3 &= \text{log time at current address}, \\ X_4 &= \text{log time in current job}, \\ e &= \text{error}. \end{aligned}$$

Credit scores

```
lm(formula = score ~ log.savings + log.income + log.address +
    log.employed, data = creditlog)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.219      5.231   -0.04   0.9667
log.savings    10.353      0.612   16.90  < 2e-16
log.income      5.052      1.258    4.02  6.8e-05
log.address     2.667      0.434    6.14  1.7e-09
log.employed    1.314      0.409    3.21   0.0014

Residual standard error: 10.2 on 495 degrees of freedom
Multiple R-squared: 0.47, Adjusted R-squared: 0.466
F-statistic:  110 on 4 and 495 DF,  p-value: <2e-16

    CV   AIC AICc   BIC   AdjR2
 104.7 2326 2326 2351 0.46582
```

Interpreting regression coefficients

- The ideal scenario is when the predictors are uncorrelated.
  - Each coefficient can be interpreted and tested separately.
- Correlations amongst predictors cause problems.
  - The variance of all coefficients tends to increase, sometimes dramatically.
  - Interpretations become hazardous – when X_j changes, everything else changes.
  - Predictions still work provided new X values are within the range of training X values.
- Claims of causality should be avoided for observational data.

Interactions

- An interaction occurs when the one variable changes the effect of a second variable. (e.g., spending on radio advertising increases the effectiveness of TV advertising).
- To model an interaction, include the product X_1X_2 in the model in addition to X_1 and X_2.
- **Hierarchy principle**: If we include an interaction in a model, we should also include the main effects, even if the p-values associated with their coefficients are not significant. (This is because the interactions are almost impossible to interpret without the main effects.)

Residual patterns

- If a plot of the residuals vs any predictor in the model shows a pattern, then the relationship is nonlinear.
- If a plot of the residuals vs any predictor **not** in the model shows a pattern, then the predictor should be added to the model.
- If a plot of the residuals vs fitted values shows a pattern, then there is heteroscedasticity in the errors. (Could try a transformation.)

Residual patterns

- If a plot of the residuals vs any predictor in the model shows a pattern, then the relationship is nonlinear.
- If a plot of the residuals vs any predictor **not** in the model shows a pattern, then the predictor should be added to the model.
- If a plot of the residuals vs fitted values shows a pattern, then there is heteroscedasticity in the errors. (Could try a transformation.)

Residual patterns

- If a plot of the residuals vs any predictor in the model shows a pattern, then the relationship is nonlinear.
- If a plot of the residuals vs any predictor **not** in the model shows a pattern, then the predictor should be added to the model.
- If a plot of the residuals vs fitted values shows a pattern, then there is heteroscedasticity in the errors. (Could try a transformation.)

Outline

1 Revision: multiple regression

2 Matrix formulation

3 Splines

Matrix formulation

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{e} = (e_1, \ldots, e_n)'$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \ldots & X_{p,1} \\ 1 & X_{1,2} & X_{2,2} & \ldots & X_{p,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1,n} & X_{2,n} & \ldots & X_{p,n} \end{bmatrix}.$$

Then

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}.$$

Matrix formulation

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{e} = (e_1, \ldots, e_n)'$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \ldots & X_{p,1} \\ 1 & X_{1,2} & X_{2,2} & \ldots & X_{p,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1,n} & X_{2,n} & \ldots & X_{p,n} \end{bmatrix}.$$

Then

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}.$$

Matrix formulation

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + e_i.$$

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{e} = (e_1, \ldots, e_n)'$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \ldots & X_{p,1} \\ 1 & X_{1,2} & X_{2,2} & \ldots & X_{p,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1,n} & X_{2,n} & \ldots & X_{p,n} \end{bmatrix}.$$

Then

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}.$$

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

(The "normal equation".)

$$\hat{\sigma}^2 = \frac{1}{n-k-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$$

Note: If you fall for the dummy variable trap, $(\boldsymbol{X}'\boldsymbol{X})$ is a singular matrix.

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

(The “normal equation”.)

$$\hat{\sigma}^2 = \frac{1}{n-k-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$$

Note: If you fall for the dummy variable trap, $(\boldsymbol{X}'\boldsymbol{X})$ is a singular matrix.

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

(The “normal equation”.)

$$\hat{\sigma}^2 = \frac{1}{n-k-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$$

Note: If you fall for the dummy variable trap, $(\boldsymbol{X}'\boldsymbol{X})$ is a singular matrix.

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

(The “normal equation”.)

$$\hat{\sigma}^2 = \frac{1}{n-k-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$$

Note: If you fall for the dummy variable trap, $(\boldsymbol{X}'\boldsymbol{X})$ is a singular matrix.

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)\right)$$

which is maximized when $(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)$ is minimized.

So **MLE = OLS**.

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)\right)$$

which is maximized when $(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)$ is minimized.

So **MLE = OLS**.

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)\right)$$

which is maximized when $(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$ is minimized.

So **MLE = OLS**.

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)\right)$$

which is maximized when $(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$ is minimized.

So **MLE = OLS**.

Multiple regression predictions

Optimal predictions

$$\hat{Y}^* = \mathrm{E}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*) = \boldsymbol{X}^*\hat{\beta} = \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

where $\boldsymbol{X}^*$ is a row vector containing the values of the regressors for the predictions (in the same format as $\boldsymbol{X}$).

Prediction variance

$$\mathrm{Var}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*) = \sigma^2\left[1 + \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}^*)'\right]$$

- This ignores any errors in $\boldsymbol{X}^*$.
- 95% prediction intervals assuming normal errors: $\hat{Y}^* \pm 1.96\sqrt{\mathrm{Var}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*)}$.

Multiple regression predictions

Optimal predictions

$$\hat{Y}^* = \mathrm{E}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*) = \boldsymbol{X}^*\hat{\boldsymbol{\beta}} = \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

where $\boldsymbol{X}^*$ is a row vector containing the values of the regressors for the predictions (in the same format as $\boldsymbol{X}$).

Prediction variance

$$\mathrm{Var}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*) = \sigma^2\left[1 + \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}^*)'\right]$$

- This ignores any errors in $\boldsymbol{X}^*$.
- 95% prediction intervals assuming normal errors: $\hat{Y}^* \pm 1.96\sqrt{\mathrm{Var}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*)}$.

Multiple regression predictions

Optimal predictions

$$\hat{Y}^* = \mathrm{E}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*) = \boldsymbol{X}^*\hat{\boldsymbol{\beta}} = \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

where $\boldsymbol{X}^*$ is a row vector containing the values of the regressors for the predictions (in the same format as $\boldsymbol{X}$).

Prediction variance

$$\mathrm{Var}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*) = \sigma^2\left[1 + \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}^*)'\right]$$

- This ignores any errors in $\boldsymbol{X}^*$.
- 95% prediction intervals assuming normal errors: $\hat{Y}^* \pm 1.96\sqrt{\mathrm{Var}(Y^*|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{X}^*)}$.

Multiple regression predictions

Optimal predictions

$$\hat{Y}^* = \mathrm{E}(Y^*|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{X}^*) = \boldsymbol{X}^*\hat{\beta} = \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

where $\boldsymbol{X}^*$ is a row vector containing the values of the regressors for the predictions (in the same format as $\boldsymbol{X}$).

Prediction variance

$$\mathrm{Var}(Y^*|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{X}^*) = \sigma^2\left[1 + \boldsymbol{X}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}^*)'\right]$$

- This ignores any errors in $\boldsymbol{X}^*$.
- 95% prediction intervals assuming normal errors: $\hat{Y}^* \pm 1.96\sqrt{\mathrm{Var}(Y^*|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{X}^*)}$.

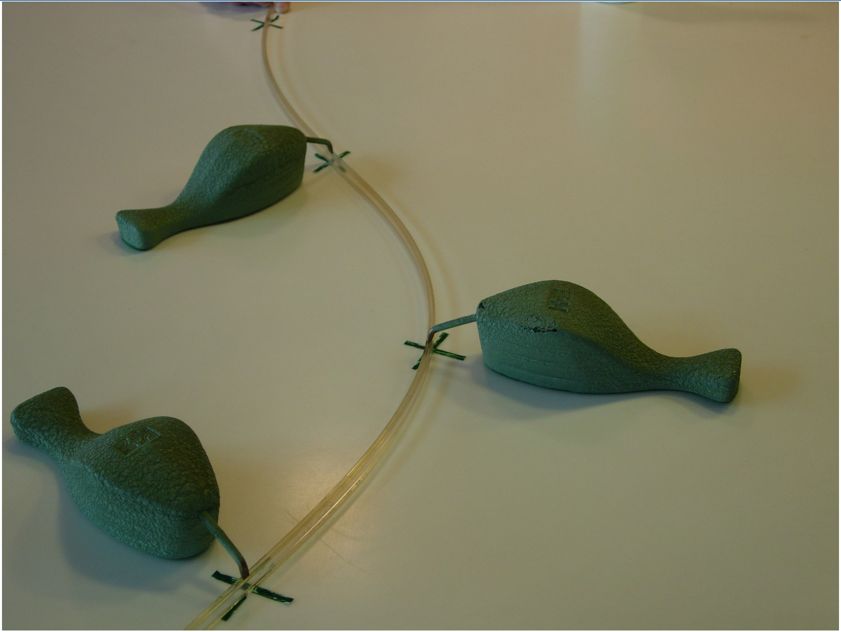
Outline

1 Revision: multiple regression

2 Matrix formulation

3 Splines

Splines



Splines

Splines

Splines

Knots: $\kappa_1, \ldots, \kappa_K$.

A spline is a continuous function $f(x)$ consisting of polynomials between each consecutive pair of 'knots' $x = \kappa_j$ and $x = \kappa_{j+1}$.

- Parameters constrained so that $f(x)$ is continuous.
- Further constraints imposed to give continuous derivatives.

Splines

Knots: $\kappa_1, \ldots, \kappa_K$.

A spline is a continuous function $f(x)$ consisting of polynomials between each consecutive pair of 'knots' $x = \kappa_j$ and $x = \kappa_{j+1}$.

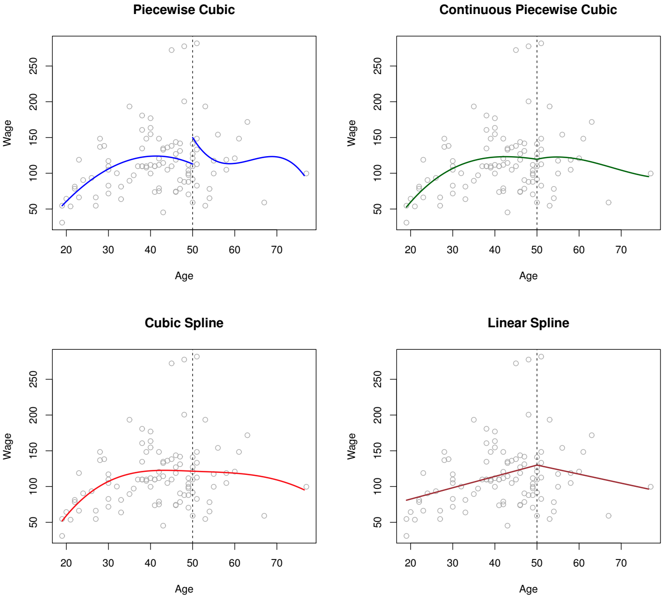
- Parameters constrained so that $f(x)$ is continuous.
- Further constraints imposed to give continuous derivatives.

Splines

Knots: $\kappa_1, \ldots, \kappa_K$.

A spline is a continuous function $f(x)$ consisting of polynomials between each consecutive pair of 'knots' $x = \kappa_j$ and $x = \kappa_{j+1}$.

- Parameters constrained so that $f(x)$ is continuous.
- Further constraints imposed to give continuous derivatives.


Piecewise Cubic
Continuous Piecewise Cubic
Cubic Spline
Linear Spline
Wage
Age
50
100
150
200
250
20
30
40
50
60
70

Truncated power basis

- Predictors: $x, \ldots, x^p, (x-\kappa_1)_+^p, \ldots, (x-\kappa_K)_+^p$
- Then the regression is piecewise order-p polynomials.
- $p-1$ continuous derivatives.
- Usually choose $p=1$ or $p=3$.
- $p+K+1$ degrees of freedom

Truncated power basis

- Predictors: $x, \ldots, x^p, (x - \kappa_1)_+^p, \ldots, (x - \kappa_K)_+^p$
- Then the regression is piecewise order-p polynomials.
- $p - 1$ continuous derivatives.
- Usually choose $p = 1$ or $p = 3$.
- $p + K + 1$ degrees of freedom

Truncated power basis

- Predictors: $x, \ldots, x^p, (x - \kappa_1)_+^p, \ldots, (x - \kappa_K)_+^p$
- Then the regression is piecewise order-p polynomials.
- $p - 1$ continuous derivatives.
- Usually choose $p = 1$ or $p = 3$.
- $p + K + 1$ degrees of freedom

Truncated power basis

- Predictors: $x, \dots, x^p, (x - \kappa_1)^p_+, \dots, (x - \kappa_K)^p_+$
- Then the regression is piecewise order-p polynomials.
- $p - 1$ continuous derivatives.
- Usually choose $p = 1$ or $p = 3$.
- $p + K + 1$ degrees of freedom

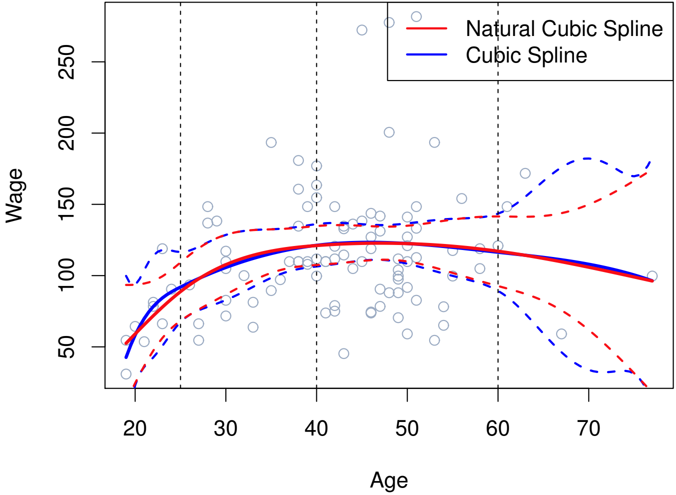
Truncated power basis

- Predictors: $x, \ldots, x^p, (x - \kappa_1)_+^p, \ldots, (x - \kappa_K)_+^p$
- Then the regression is piecewise order-p polynomials.
- $p - 1$ continuous derivatives.
- Usually choose $p = 1$ or $p = 3$.
- $p + K + 1$ degrees of freedom

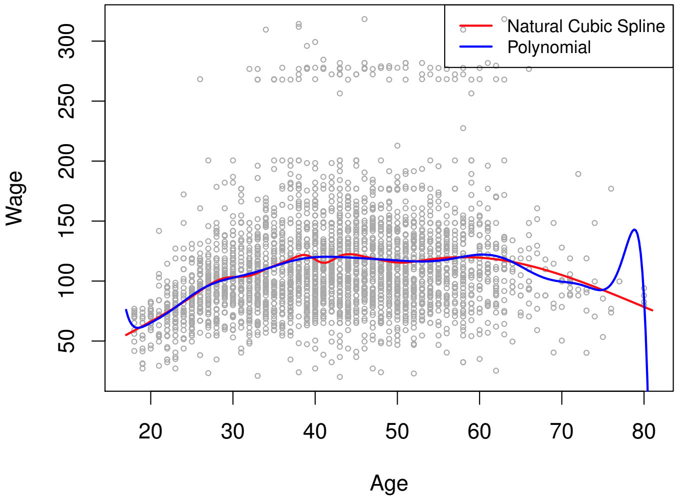
Natural splines

- Splines based on truncated power bases have high variance at the outer range of the predictors.
- Natural splines are similar, but have additional boundary constraints: the function is linear at the boundaries. This reduces the variance.
- Degrees of freedom $\mathrm{df} = K + 1$ (i.e., there are $K = \mathrm{df} - 1$ interior knots.)
- Create predictors using ns function in R (automatically chooses knots given df).

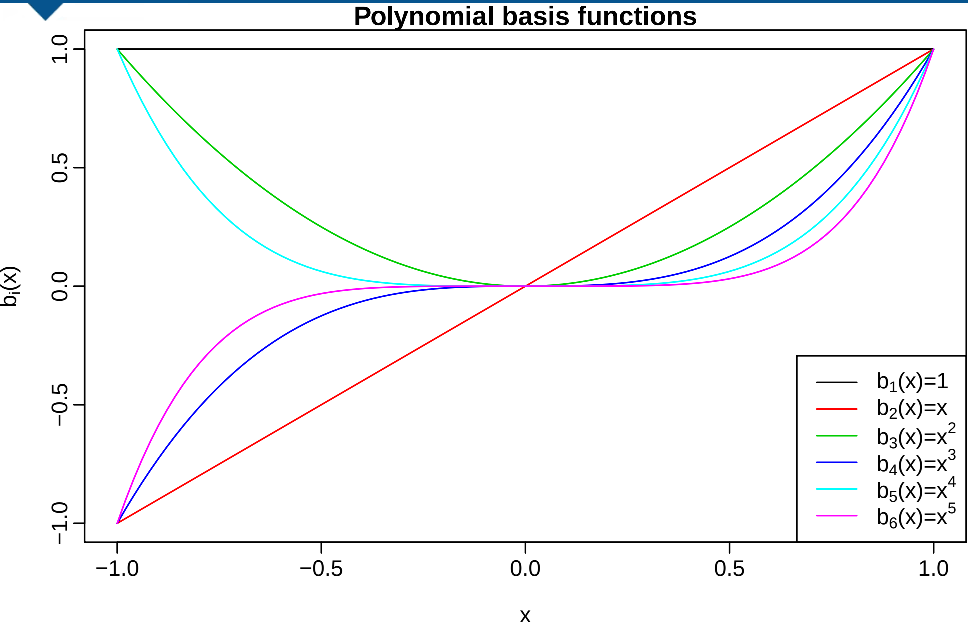
Natural splines



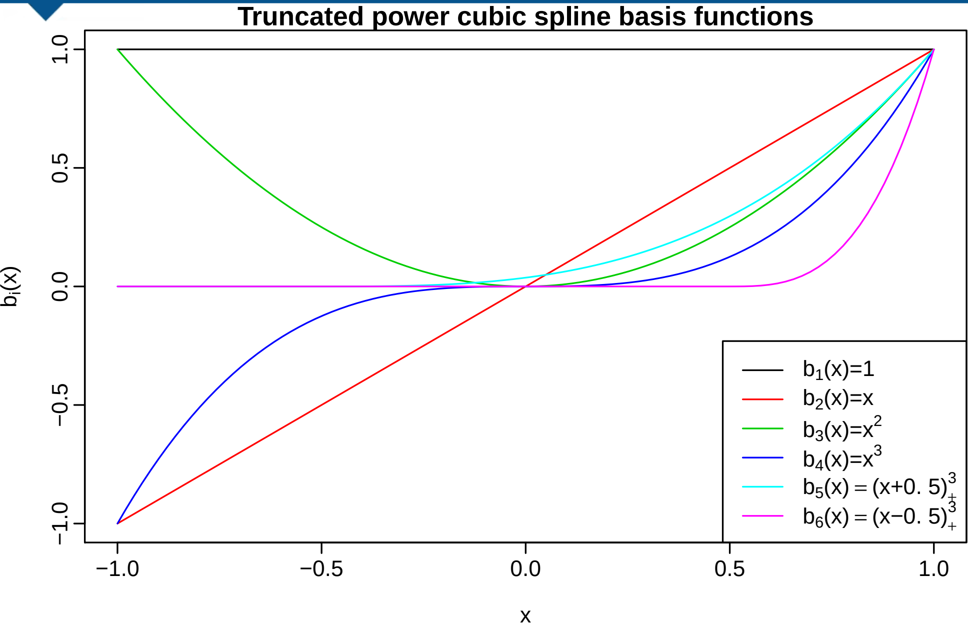
Natural splines



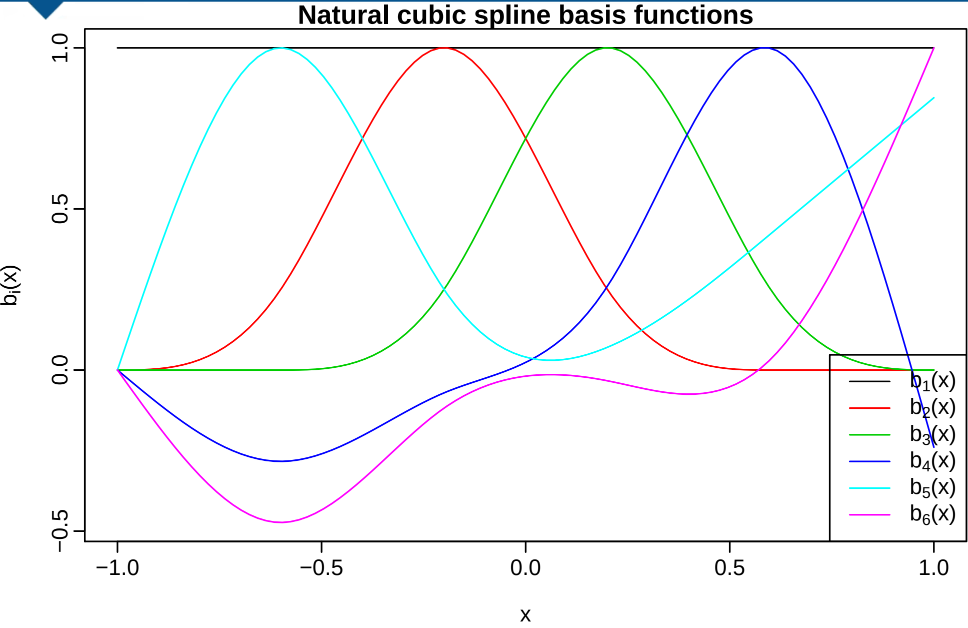
Basis functions



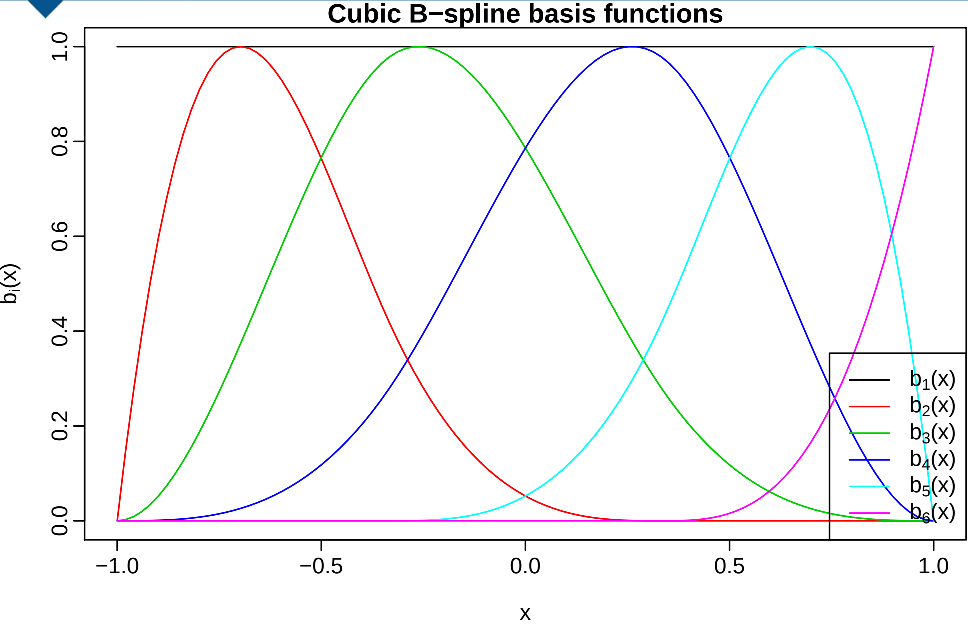
Basis functions



Basis functions



Basis functions



Nonlinear choices

1. Polynomials (beware)
2. Truncated power basis splines
3. Natural splines
4. B-splines
5. Smoothing splines
6. Kernel regression
7. Local regression
8. kNN