

Recap
Lecture 3

Linear Regression II

ISLR 3, ESL 3



Krikamol Muandet
Jilles Vreeken



UNIVERSITÄT
DES
SAARLANDES



CISPA
HELMHOLTZ CENTER FOR
INFORMATION SECURITY

Recap (Lecture 3)

- Usefulness of Predictors
- Beyond Additivity
- Regression Pitfalls
- kNN Regression

Recap (Lecture 3)

- Usefulness of Predictors
 - Is at least one predictor useful?
 - Which subset of predictors are useful? (preview, covered in a coming lecture)

Usefulness of predictors: any?

	Coefficient	Std. error	-statistic	-value
intercept	2.939	0.3119	9.42	<0.0001
TV	0.046	0.0014	32.81	<0.0001
radio	0.189	0.0086	21.89	<0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

$\beta \gg 0 \rightarrow$ Potentially useful predictor

$\beta \approx 0 \rightarrow$ Probably not a useful predictor

Usefulness of predictors: any?

	Coefficient	Std. error	-statistic	-value
intercept	2.939	0.3119	9.42	<0.0001
TV	0.046	0.0014	32.81	<0.0001
radio	0.189	0.0086	21.89	<0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

$\beta \gg 0 \rightarrow$ Potentially useful predictor

$\beta \approx 0 \rightarrow$ Probably not a useful predictor

- Hypothesis Testing: is there at least one predictor with $\beta \neq 0$?

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad \text{vs.} \quad H_a: \text{at least one } \beta_i \text{ is non-zero}$$

Usefulness of predictors: which?

	Coefficient	Std. error	-statistic	-value
intercept	2.939	0.3119	9.42	<0.0001
TV	0.046	0.0014	32.81	<0.0001
radio	0.189	0.0086	21.89	<0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

{}, {**TV**},{**radio**},{**newspaper**},
{**TV, radio**},{**TV, newspaper**},
{**radio, newspaper**},
{**TV, radio, newspaper**}

- Subset Selection Techniques

Recap (Lecture 3)

- Usefulness of Predictors and Model
- Beyond Additivity
 - Interactions between predictors
 - Nonlinear relationship between predictor and target (preview, covered in a coming lecture)

Interactions between predictors

	Coefficient	Std. error	-statistic	-value
intercept	6.7502	0.248	27.23	<0.0001
TV	0.0191	0.002	12.70	<0.0001
radio	0.0289	0.009	3.24	0.0014
radio × TV	0.0011	0.000	20.73	<0.0001

- Interaction terms as additional predictors

$$\begin{aligned}\mathbf{sales} &= \beta_0 + \beta_1 \times \mathbf{TV} + \beta_2 \times \mathbf{radio} + \beta_3 \times (\mathbf{radio} \times \mathbf{TV}) + \epsilon \\ &= \beta_0 + (\beta_1 + \beta_3 \times \mathbf{radio}) \times \mathbf{TV} + \beta_2 \times \mathbf{radio} + \epsilon\end{aligned}$$

Nonlinear relationship between predictor and target

	Coefficient	Std. error	-statistic	-value
intercept	56.9001	1.8004	31.6	<0.0001
horsepower	-0.4662	0.0311	-15.0	<0.0001
horsepower²	0.0012	0.0001	10.1	<0.0001

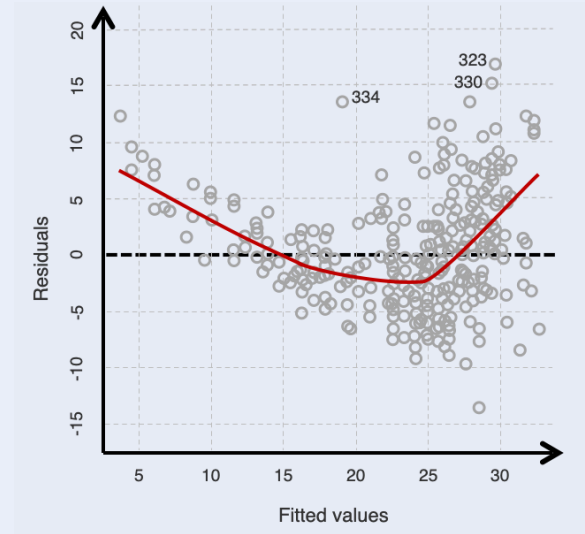
- Polynomial terms as additional predictors

$$\text{mpg} = \beta_0 + \beta_1 \times \text{horsepower} + \beta_2 \times \text{horsepower}^2 + \epsilon$$

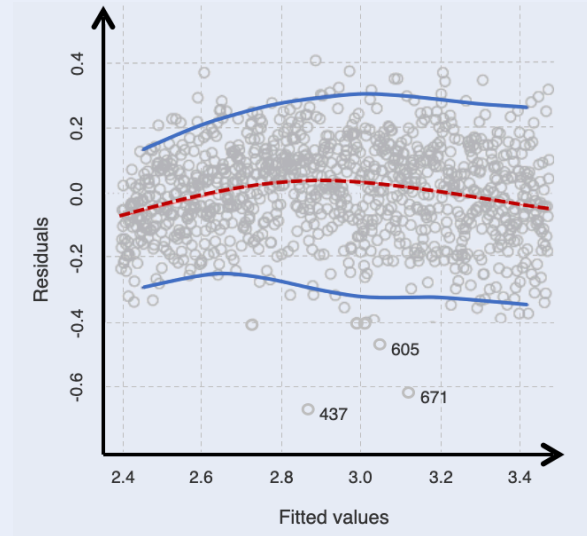
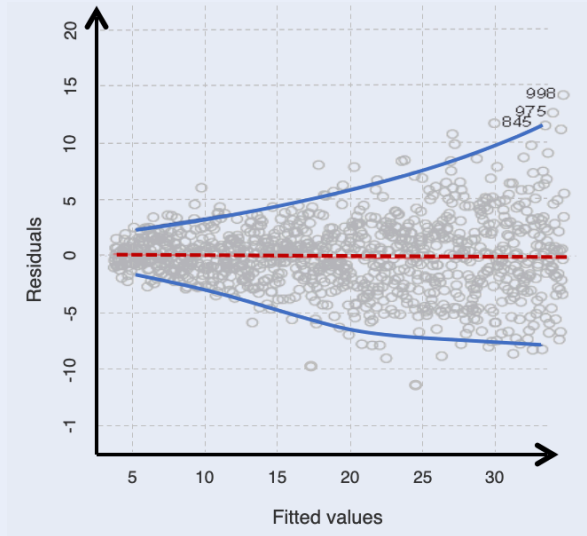
Recap (Lecture 3)

- Usefulness of Predictors and Model
- Beyond Additivity
- Regression Pitfalls
 - Nonlinearity
 - Correlated noise
 - Dependent noise (heteroskedasticity)
 - Dependent predictors (collinearity)
 - Outliers and high-leverage points

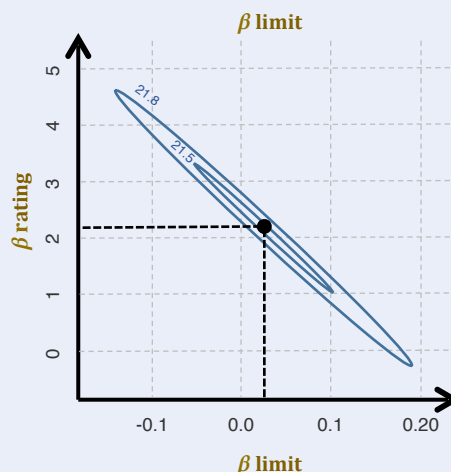
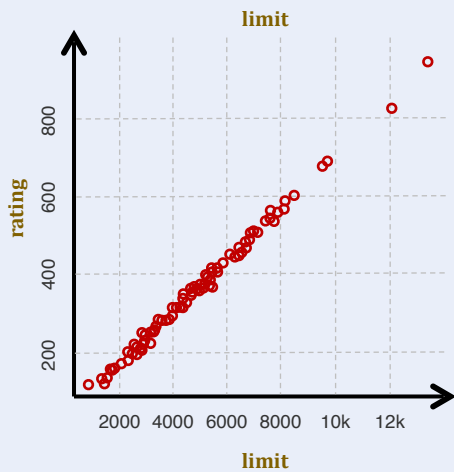
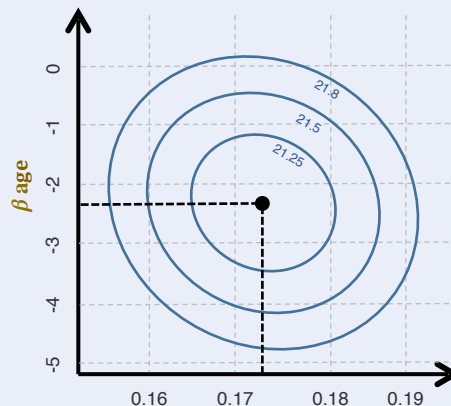
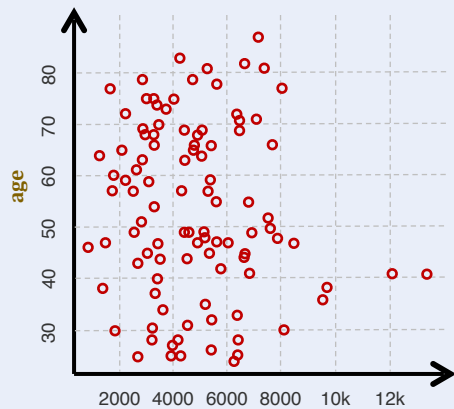
Which residual plot shows a case of nonlinearity?



What pitfall of regression is illustrated here?



Which predictors (red) and coefficients (blue) are collinear?



How do the two kNN plots differ?

