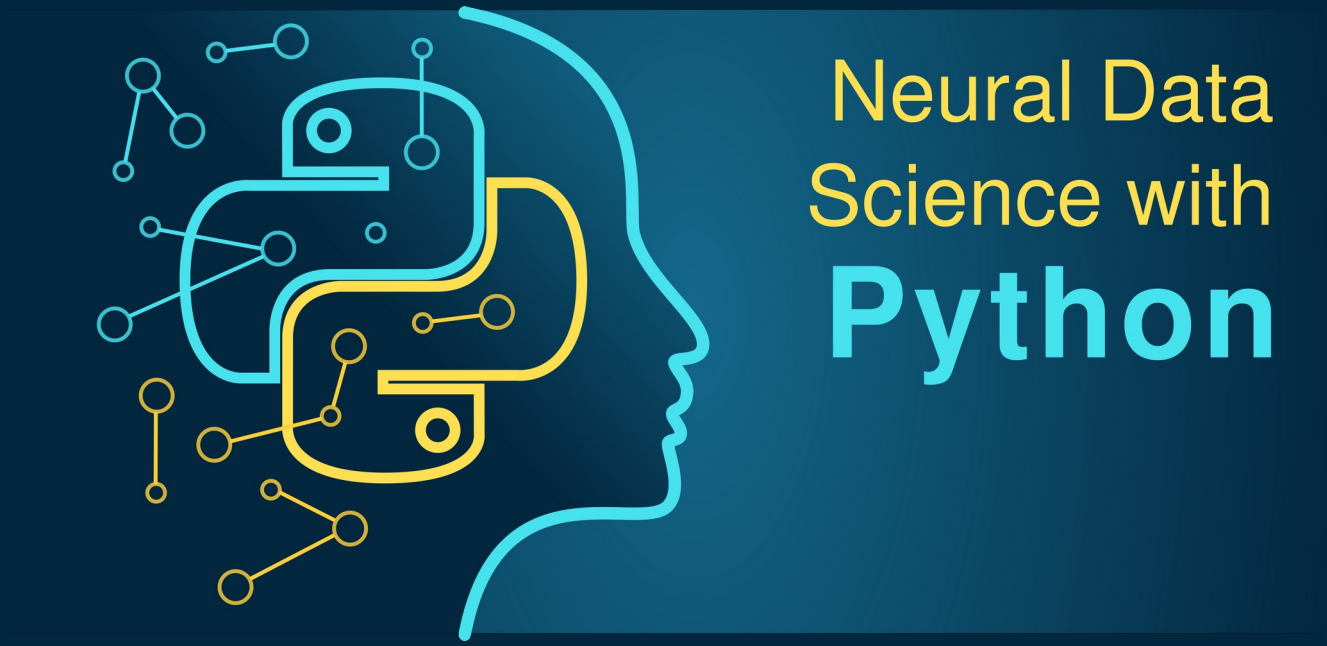


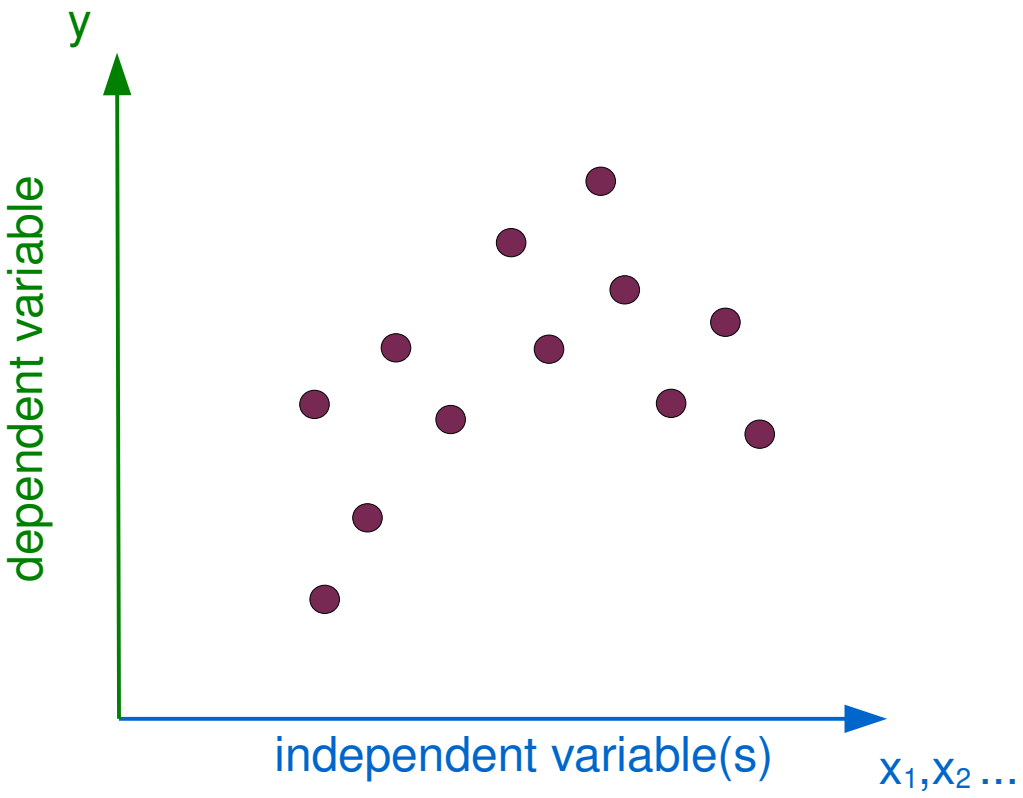
L10 : Nonlinear regression

Michael Graupner

SPPIN – Saint-Pères Institute for the Neurosciences

Université de Paris, CNRS

Reminder : regression analysis

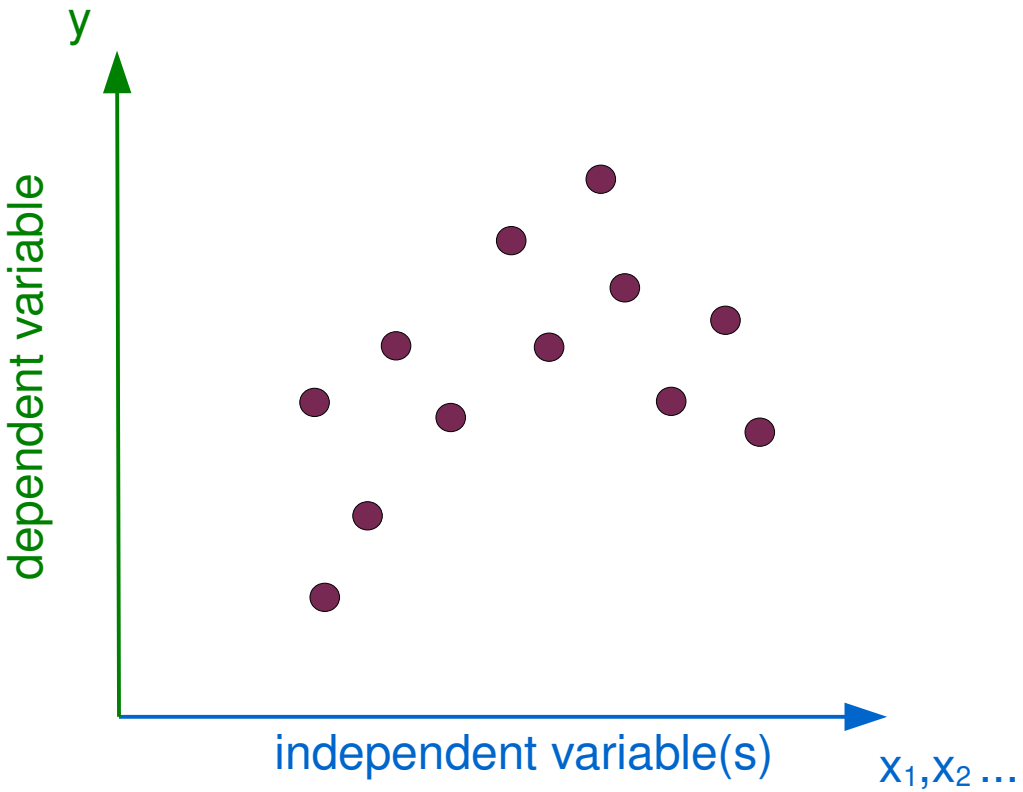


Dependent Variable: This is the main factor that we are trying to understand or predict.

Independent Variables (or predictors): These are the factors that we hypothesize have an impact on your dependent variable.

Observations: Data points → measured relations between independent and dependent variable.

Assumption : Relationship between X and Y



Regression Analysis : Approaches to examine the relationship between the variables, *i.e.*, to estimate f.

$$Y = f(X) + \epsilon$$

X ... independent variables or predictors

Y ... dependent or respons variable

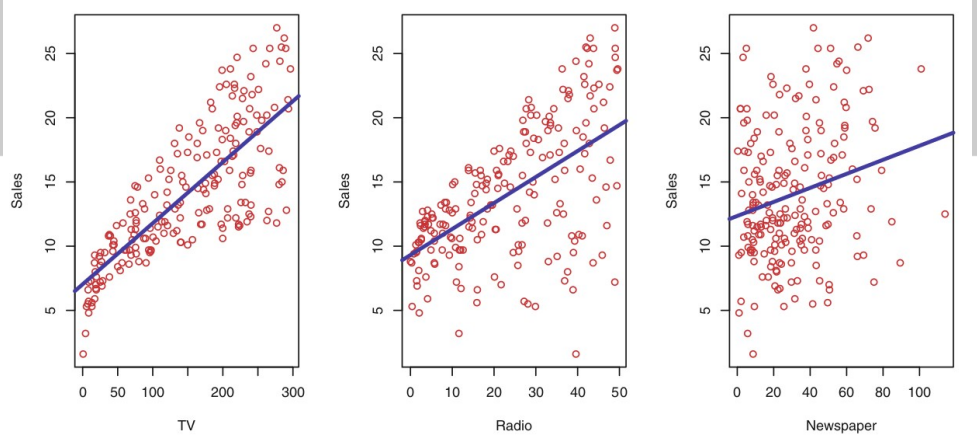
f ... fixed but unknown function of $X_1, \ldots X_p$; represents the systematic information that X provides about Y

ε ... additive error term

Why estimate *f* ? : prediction and inference

prediction	inference
X is readily available; output *Y* cannot easily be obtained : requires to predict *Y*	want to understand relationship between X an Y (how *Y* changes of function of *X*); not necessarily to make predictions
exact form of *f* is not of interest; provided it yields accurate prediction of *Y*	*f* cannot be treated as black box, we need to know exact form : - Which predictors are associated with the response? - What is the relationship between each response and the predictors? - Can the relationship btw. Y and each predictor adequately summarized using a linear equation?

Examples for prediction and inference

prediction	inference
X_1 , . . . , X_p are characteristics of a patient's blood sample that can be easily measured in a lab, and Y is a variable encoding the patient's risk for a severe adverse reaction to a particular drug. Prediction stock price in the future	Advertising data set consists of the sales of a product in different markets, along with advertising budgets for the product for three different media: **TV , radio, and newspaper.** - Which media contribute to sales? - Which media generate the biggest boost in sales? 

Prediction : reducible and irreducible error

$$Y = f(X) + \epsilon \xrightarrow{\text{estimation}} \hat{Y} = \hat{f}(X)$$

$\hat{f}$ ... represents the estimate of *f*

$\hat{Y}$ ... represents the estimate of *Y*

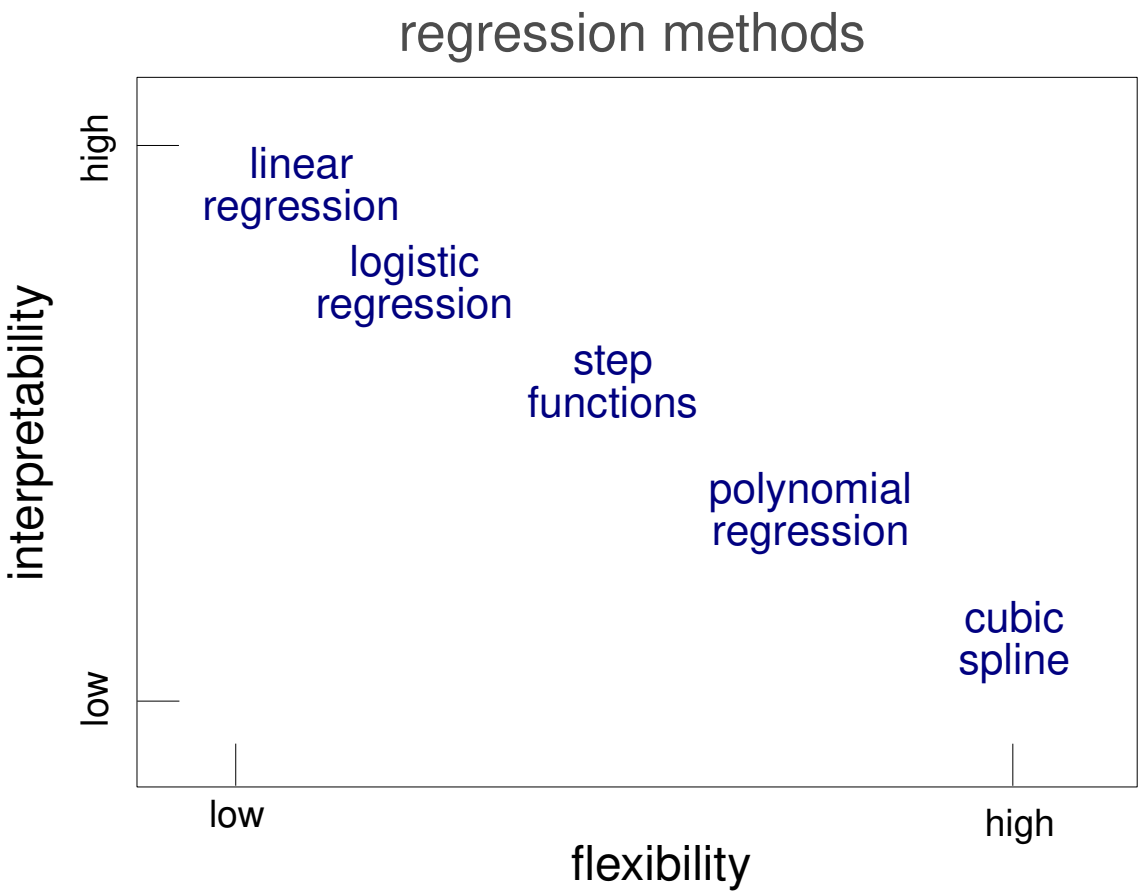
- accuracy of $\hat{Y}$ depends on the *reducible* and the *irreducible* error :
  - **reducible error :** $\hat{f}$ will not be the perfect estimate of *f*; finding the most appropriate estimate for *f* improves the accuracy of $\hat{f}$
  - **irreducible error :** even for the perfect estimate of *f*, we still have the error ϵ ; this error cannot be predicted by using *X* (e.g. unmeasured variables, unmeasurable variations)

Which method of estimating f ?

Depending on goal – prediction, inference or combination of both – different methods for estimating f might be appropriate

- **linear models :** allow simple and interpretable inference; but may not yield accurate predictions
- **highly non-linear approaches :** can provide accurate predictions of Y ; less interpretable model for which inference is challenging

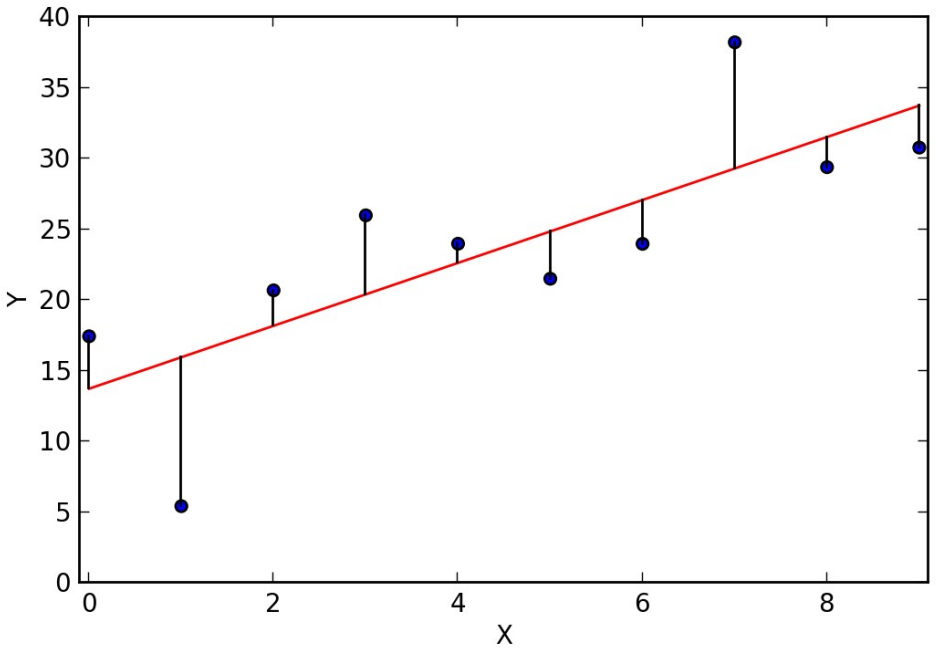
Trade-off: prediction accuracy vs model intepretability



- less flexible : produce small range of shapes to estimate *f* (e.g. linear regression → lines,planes)
- more flexible : can generate much wider range of shapes to estimate *f* (e.g. splines)

Measuring quality of fit : mean squared error (MSE)

Example : MSE for linear regression

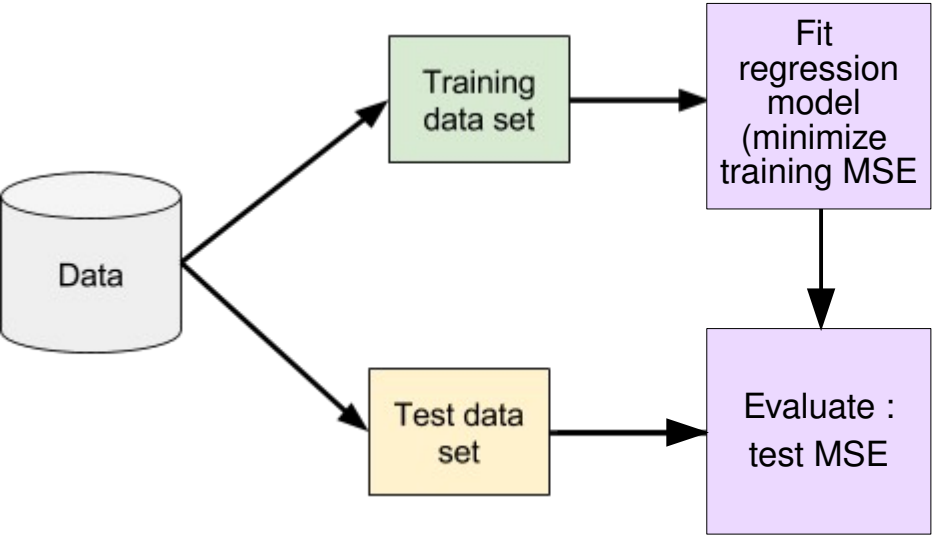


- quantifies the extent to which predicted response value is close to the true response

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}\left(x_i\right)\right)^2$$

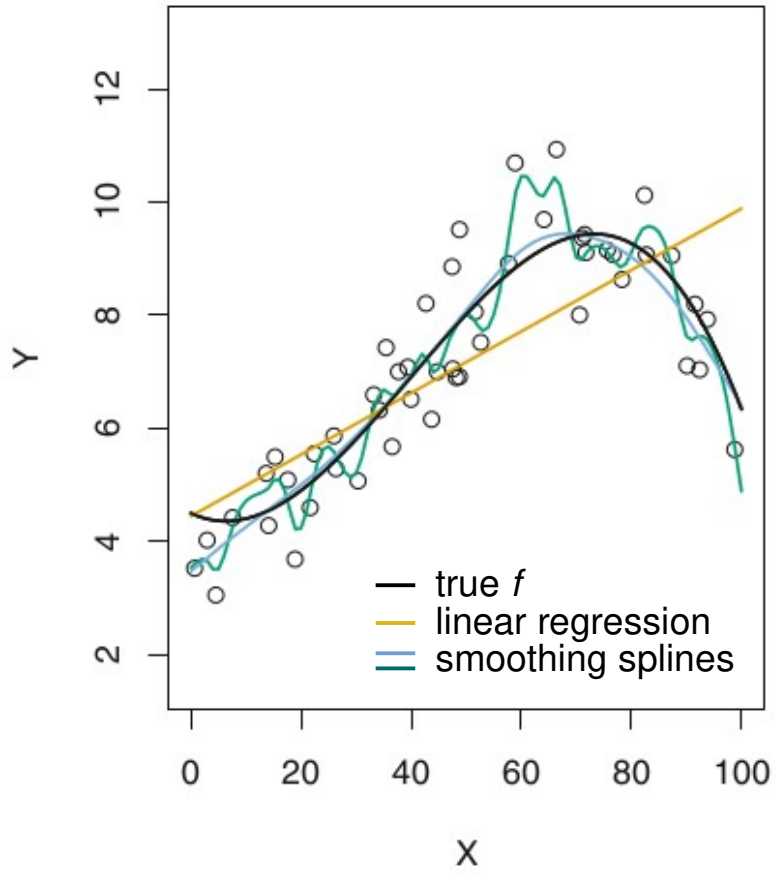
$\hat{f}\left(x_i\right)$ ... prediction that $\hat{f}$ gives for the *i*th observation

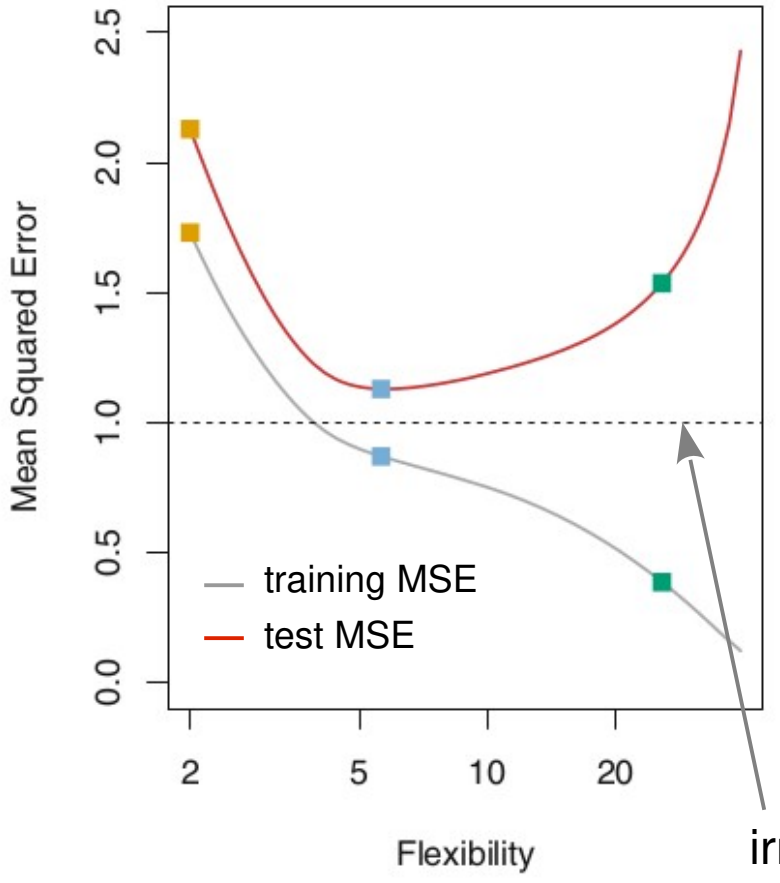
Training vs. Test mean-squared error (MSE)



- MSE computed using the training data is used to fit the model : *training MSE*
- we are interested in the accuracy of the prediction when model is applied to previously unseen test data
- want to chose the model that gives lowest *test MSE* , i.e., the MSE calculated on the previously unseen test data (as opposed to the model with lowest training MSE)
- **Attention :** model with the lowest training MSE is not necessarily the model with the lowest test MSE

Overfitting : small training MSE & large test MSE





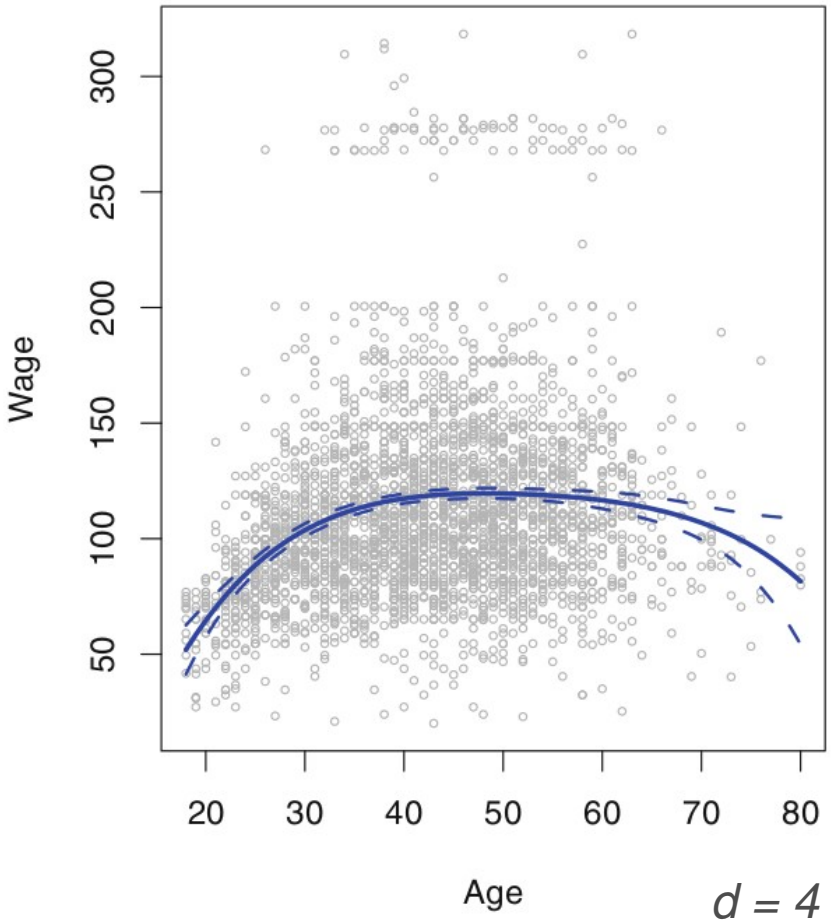
overfitting data : small training MSE and large test MSE; training to specific (random) pattern in training data which does not reflect true property of *f*

irreducible error ϵ

Regression methods : beyond linearity

- *Polynomial regression* extends the linear model by adding extra predictors, obtained by raising each of the original predictors to a power.
- *Step functions* cut the range of a variable into K distinct regions in order to produce a qualitative variable. This has the effect of fitting a piecewise constant function.
- *Regression splines* are more flexible than polynomials and step functions, and in fact are an extension of the two. They involve dividing the range of X into K distinct regions. Within each region, a polynomial function is fit to the data. However, these polynomials are constrained so that they join smoothly at the region boundaries, or knots. Provided that the interval is divided into enough regions, this can produce an extremely flexible fit.

Polynomial Regression



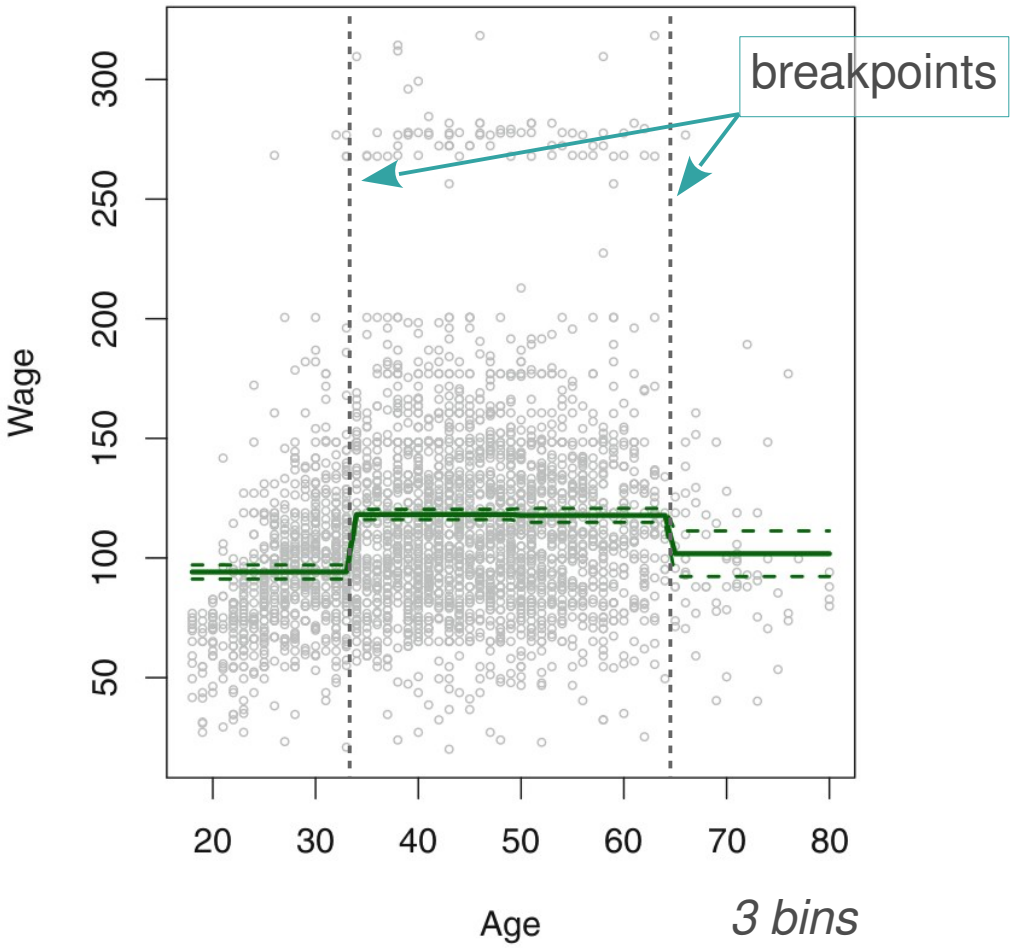
- replaces the linear model with a polynomial function

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots \beta_d X^d + \epsilon$$

- degree *d* controls the non-linearity of the curve
- unusual to use d greater than 3 or 4 : curve becomes very flexible for d>4 and take strange shapes

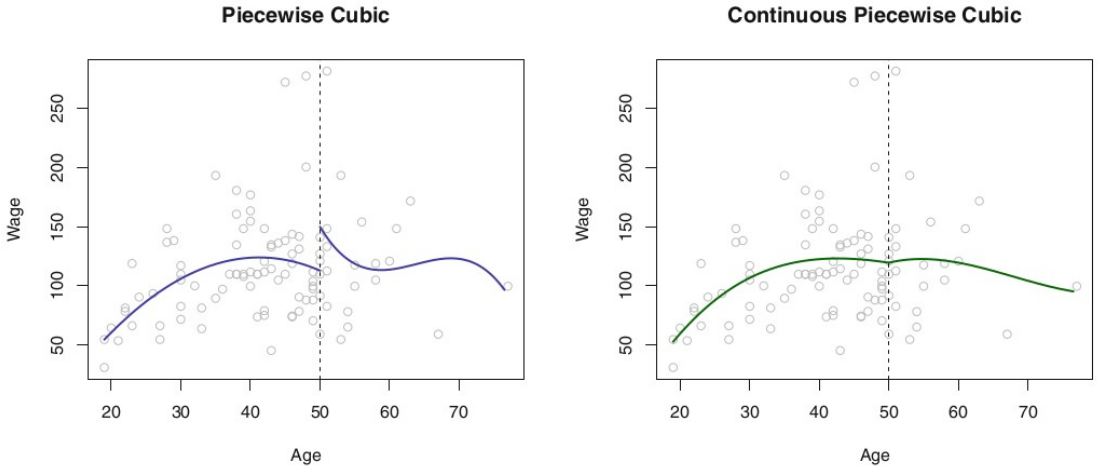
d = 4 (dashed curve 95 % confidence interval)

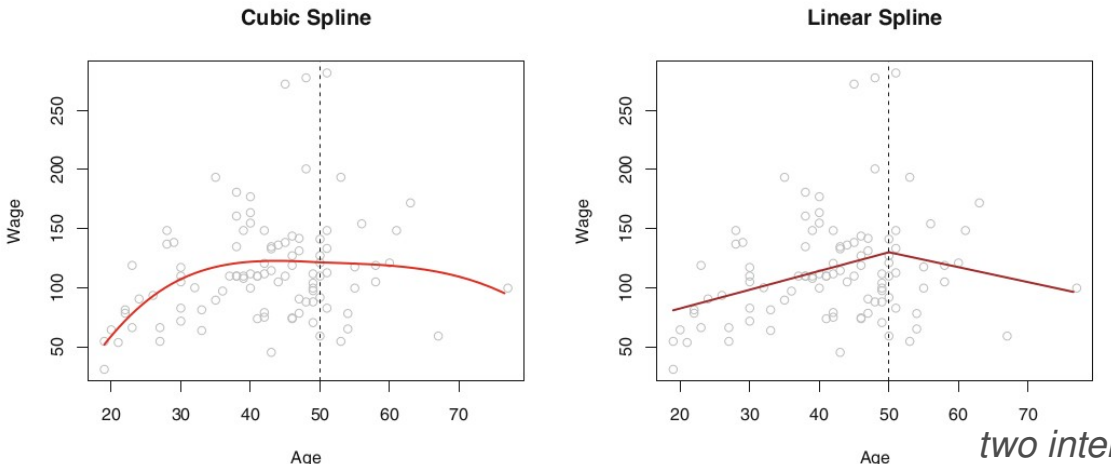
Step Functions



- break the range of X into bins and fit different constants in each bin
- breakpoints have to be defined before fitting the constants (e.g. based on percentiles)
- unless there are natural breakpoints, piecewise-constant functions can miss the action
- popular in biostatistics and epidemiology

Regression splines

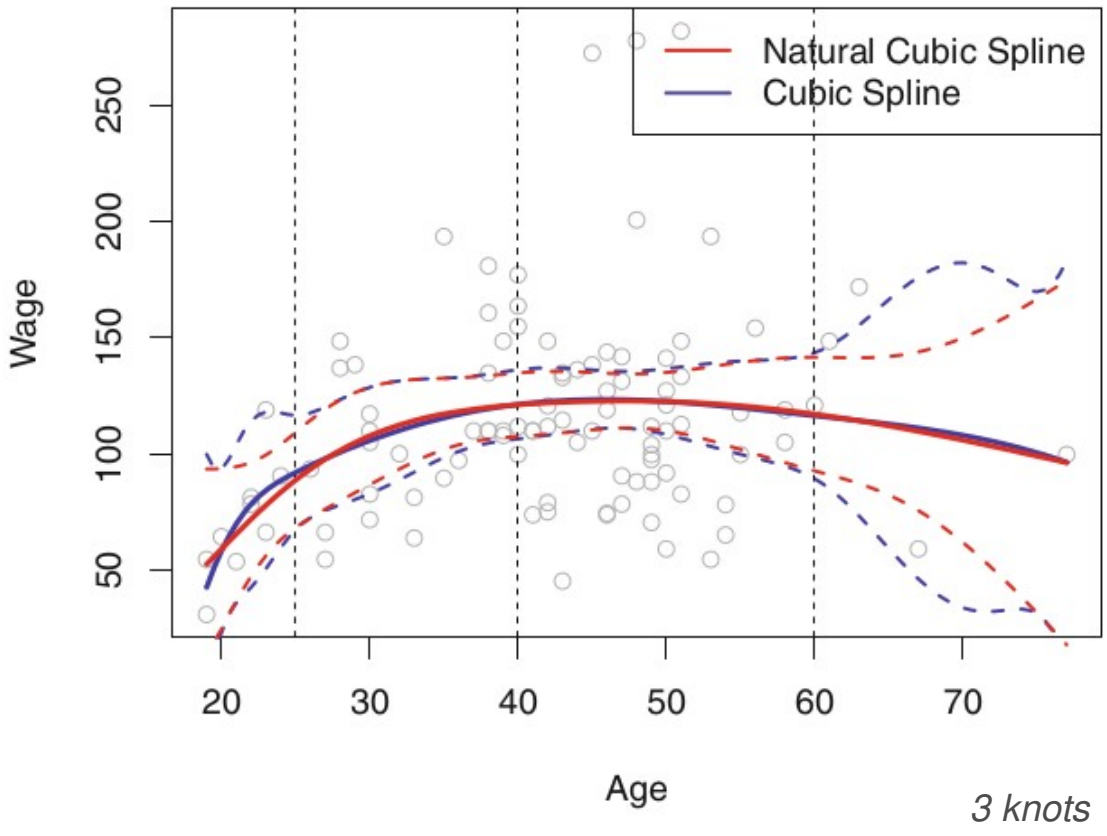




- extends polynomial regression and step function : fitting separate low-degree polynomials over different regions of *X*
- *additional constraints* are that fitted curve must be *continuous* and *smooth*
- *general definition* : degree-d spline is a piecewise degree-d polynomial with continuity in derivatives up to degree d-1

two intervals, or 1 knot

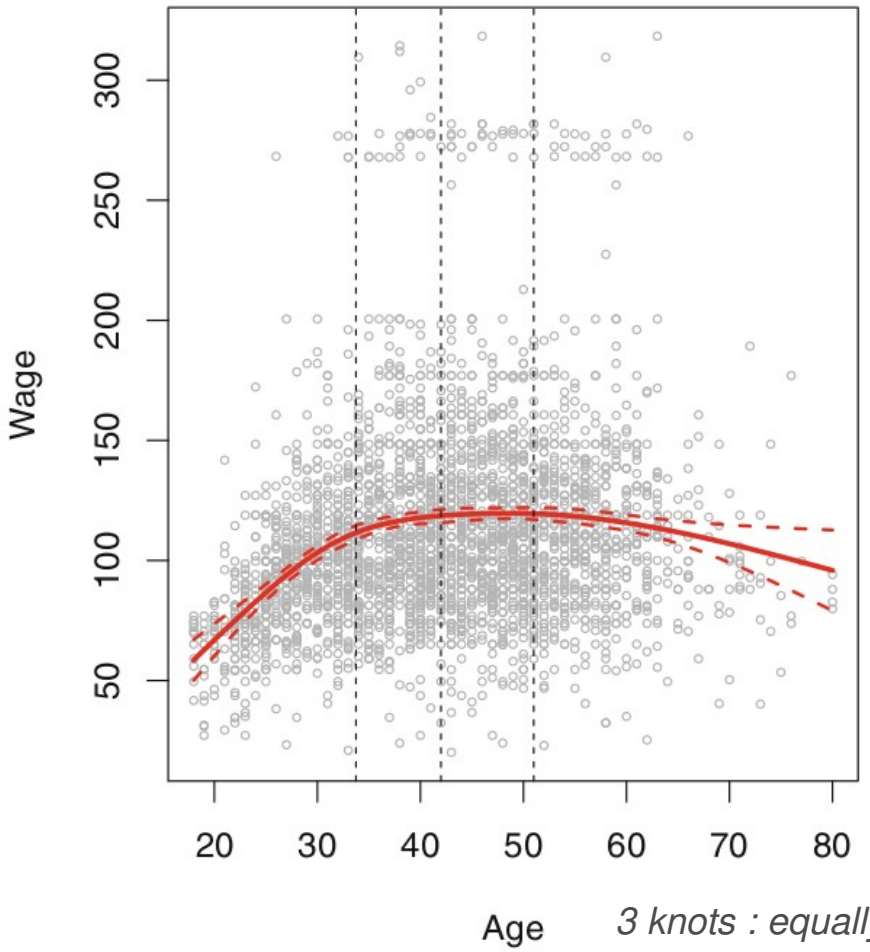
Cubic Spline and Natural Cubic Spline



3 knots

- *Cubic Spline* : fitting a cubic function (up to X^3) to the intervals and imposing continuity in 1st and 2nd derivative
- Splines have high variance at the outer range of the predictors (when X very large or small)
- Natural Spline : is a spline with additional boundary constrains – the function is required to be linear at the boundary → more stable estimates at boundaries

Choosing number and location of knots



- Where should the knots be placed ?
- *Often* : number of segments are specified and knots are placed at uniform quantiles of the data (e.g. 3 knots → knots at 25th, 50th and 75th percentile
- *Best practice* : number of knots determined through cross-validation; chosen based on smallest test RMS

3 knots : equally spaced in terms of percentiles