Machine Learning for Official Statistics and SDGs

First Live Lecture (Webinar):

Starts in **15** minutes

Christophe Bontemps, UN SIAP

UNITED NATIONS
SIAP
Statistical Institute for
Asia and the Pacific

Machine Learning for Official Statistics and SDGs

First Live Lecture (Webinar):

Starts in **10** minutes

Christophe Bontemps, UN SIAP

UNITED NATIONS
SIAP
Statistical Institute for Asia and the Pacific

Machine Learning for Official Statistics and SDGs

First Live Lecture (Webinar):

Starts in **5** minutes

Christophe Bontemps, UN SIAP

UNITED NATIONS
SIAP
Statistical Institute for Asia and the Pacific

Machine Learning for Official Statistics and SDGs

Statistical learning: *vs* Machine Learning

UNITED NATIONS
SIAP
Statistical Institute for Asia and the Pacific

[- Reminder -]

- Mute yourself always!

[- Reminder -]

- Mute yourself always!
- The lecture is recorded

[- Reminder -]

- Mute yourself always!
- The lecture is recorded
- Ask questions in the chat

[- AGENDA -]

- Introduction

[- Agenda -]

- Introduction
- Statistical learning *vs* Machine Learning

[- Agenda -]

- Introduction
- Statistical learning *vs* Machine Learning
- Q&A

[- AGENDA -]

- Introduction
- Statistical learning *vs* Machine Learning
- Q&A
- Next week

What Is Statistical Learning?

What Is Statistical Learning?

“ Statistical learning refers to a vast set of tools for understanding data”

Gareth James, Daniela Witten, Trevor Hastie , Robert Tibshirani (2021)

What Is Statistical Learning?

" Statistical learning refers to a vast set of tools for understanding data"

Gareth James, Daniela Witten, Trevor Hastie , Robert Tibshirani (2021)

What Is Statistical Learning?

" Statistical learning refers to a vast set of tools for understanding data"

Gareth James, Daniela Witten, Trevor Hastie , Robert Tibshirani (2021)

$\hookrightarrow$ Involves building statistical models

What Is Statistical Learning?

“ Statistical learning refers to a vast set of tools for understanding data”

Gareth James, Daniela Witten, Trevor Hastie , Robert Tibshirani (2021)

$\hookrightarrow$ Involves building statistical models

$\hookrightarrow$ Goals are estimation or prediction

WHAT IS STATISTICAL LEARNING?

" Statistical learning refers to a vast set of tools for understanding data"

Gareth James, Daniela Witten, Trevor Hastie , Robert Tibshirani (2021)

$\hookrightarrow$ Involves building statistical models

$\hookrightarrow$ Goals are **estimation** or prediction

What Is Statistical Learning?

" Statistical learning refers to a vast set of tools for understanding data"

Gareth James, Daniela Witten, Trevor Hastie , Robert Tibshirani (2021)

$\hookrightarrow$ Involves building statistical models
$\hookrightarrow$ Goals are estimation or **prediction**

WHAT IS STATISTICAL LEARNING?

Two main learning problems:

WHAT IS STATISTICAL LEARNING?

Two main learning problems:

- We observe **both** an *outcome* y and *explanatory* variables xs

WHAT IS STATISTICAL LEARNING?

Two main learning problems:

- We observe **both** an *outcome* y and *explanatory* variables xs

$\hookrightarrow$ **Supervised** learning

What Is Statistical Learning?

Two main learning problems:

- We observe **both** an *outcome* y and *explanatory* variables xs

$\hookrightarrow$ **Supervised** learning

Most of the examples and applications are supervised learning

What Is Statistical Learning?

Two main learning problems:

- We observe **both** an *outcome* y and *explanatory* variables xs

↪ **Supervised** learning

Most of the examples and applications are supervised learning

- We **do not** observe an outcome y but **only** several xs

What Is Statistical Learning?

Two main learning problems:

- We observe **both** an *outcome* y and *explanatory* variables xs

↪ **Supervised** learning

Most of the examples and applications are supervised learning

- We **do not** observe an outcome y but **only** several xs

↪ **Unsupervised** learning (or *cluster analysis*)

What Is Statistical Learning?

Two main learning problems:

- We observe **both** an *outcome* y and *explanatory* variables xs

↪ **Supervised** learning

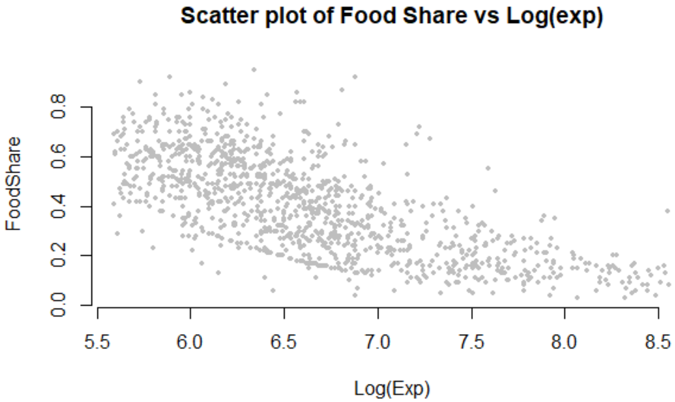
Most of the examples and applications are supervised learning

- We **do not** observe an outcome y but **only** several xs

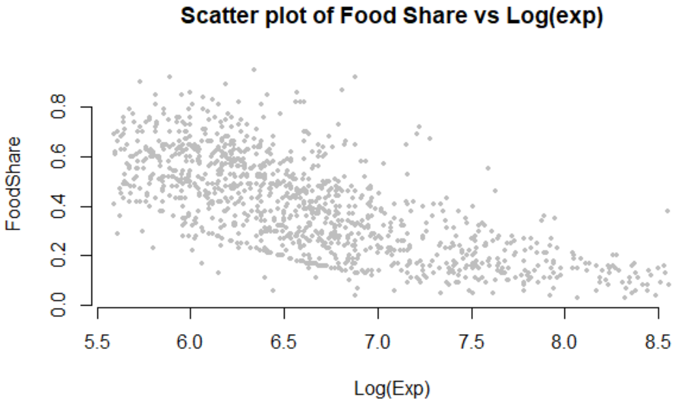
↪ **Unsupervised** learning (or *cluster analysis*)

More complex models we'll see at the end of the course

Statistical learning on an example

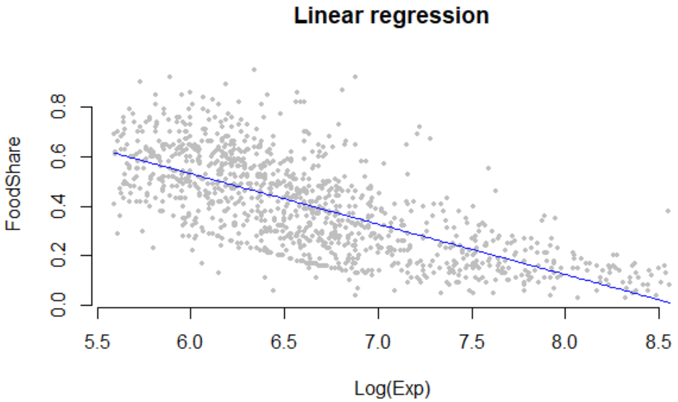


Statistical learning on an example

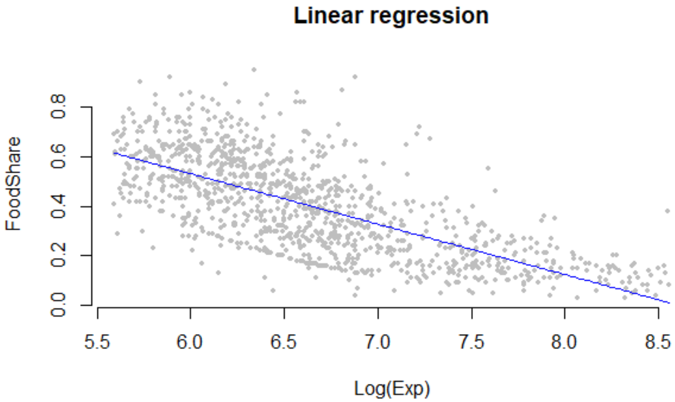


We may be interested in the **relationship** between the two variables

UNDERSTANDING = ESTIMATE $f(\cdot)$



UNDERSTANDING = ESTIMATE $f(\cdot)$



$f(\cdot)$ is the regression line

WHY ESTIMATING $f(\cdot)$?

- Inference

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y

  Identify "important" variables to understand Y

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y

  Identify "important" variables to understand Y
- Prediction

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y
  Identify "important" variables to understand Y
- Prediction

  Predict y for any **new** x using $f(\cdot)$

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y

  Identify "important" variables to understand Y
- Prediction

  Predict y for any **new** x using $f(\cdot)$
- In practice we must estimate $f(\cdot)$ using a model:

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y
  Identify "important" variables to understand Y
- Prediction

  Predict y for any **new** x using $f(\cdot)$
- In practice we must estimate $f(\cdot)$ using a model:

$$y = f(x) + \varepsilon$$

WHY ESTIMATING $f(\cdot)$?

- Inference

  Understand the nature of the relationship between X and Y
  Identify "important" variables to understand Y
- Prediction

  Predict y for any **new** x using $f(\cdot)$
- In practice we must estimate $f(\cdot)$ using a model:

$$y = f(x) + \varepsilon$$

We denote by $\widehat{f(\cdot)}$ the estimate of $f(\cdot)$

HOW TO ESTIMATE $f(\cdot)$?

- Parametric methods

How to estimate $f(\cdot)$?

- Parametric methods

  Specify a form for $f(\cdot)$, for example linear:

HOW TO ESTIMATE $f(\cdot)$?

- Parametric methods

  Specify a form for $f(\cdot)$, for example linear:

$$y = \beta_0 + \beta_1 x_+ \varepsilon$$

How to estimate $f(\cdot)$?

- Parametric methods

  Specify a form for $f(\cdot)$, for example linear:

$$y = \beta_0 + \beta_1 x_+ \varepsilon$$

- The goal is to find the line that is **minimizing** the distance to the observed points (x_i, y_i). The distance is computed as the Mean Square Error (MSE):

$$MSE(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} \big(y_i - (\beta_0 + \beta_1 x_i)\big)^2$$

How to estimate $f(\cdot)$?

- Parametric methods

  Specify a form for $f(\cdot)$, for example linear:

$$y = \beta_0 + \beta_1 x_+ \varepsilon$$

- The goal is to find the line that is **minimizing** the distance to the observed points (x_i, y_i). The distance is computed as the Mean Square Error (MSE):

$$MSE(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

- The regression line, defined by β_0 and β_1, is simply the solution of:

$$Min_{\ (\beta_0, \beta_1)}\ MSE(\beta_0, \beta_1)$$

HOW TO ESTIMATE $f(\cdot)$?

- Parametric methods

  Specify a form for $f(\cdot)$, for example linear:

$$y = \beta_0 + \beta_1 x_+ \varepsilon$$

  - The goal is to find the line that is **minimizing** the distance to the observed points (x_i, y_i). The distance is computed as the Mean Square Error (MSE):

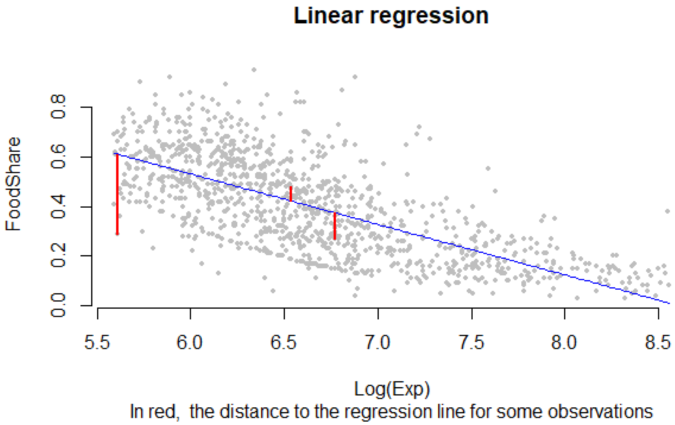
$$MSE(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

  - The regression line, defined by β_0 and β_1, is simply the solution of:

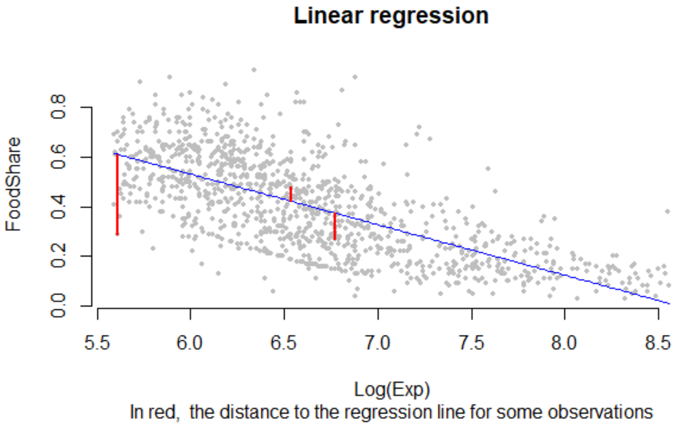
$$Min_{\ (\beta_0, \beta_1)}\ MSE(\beta_0, \beta_1)$$

The MSE it is the *cost function* minimized to determine $(\widehat{\beta}_0, \widehat{\beta}_1)$

How to estimate $f(\cdot)$: In practice



HOW TO ESTIMATE $f(\cdot)$: IN PRACTICE



The regression line is found by minimizing the sum of all distances or **MSE**

RESULTS: $\widehat{f(\cdot)}$

From the result and the estimated parameters $(\widehat{\beta}_0, \widehat{\beta}_1)$, we see that there is a relation, and that it is decreasing.

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	1.75	0.04	41.09	0
ltexp	-0.20^{***}	0.01	-31.84	0

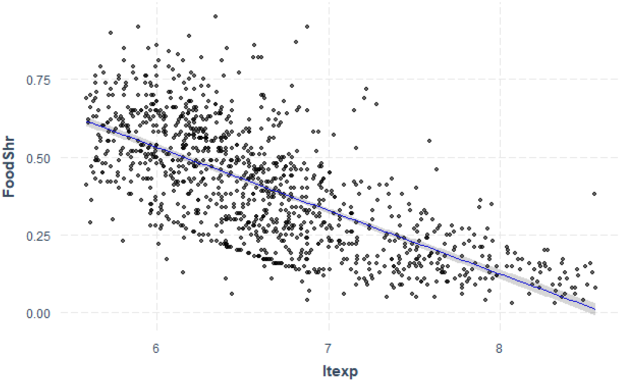
The quality of the adjustment may be measured by the $R^2 = 0.478$

Beyond linearity

Beyond linearity

- A linear model may be unadapted or too simple

$$y = \beta_0 + \beta_1 x + \varepsilon$$

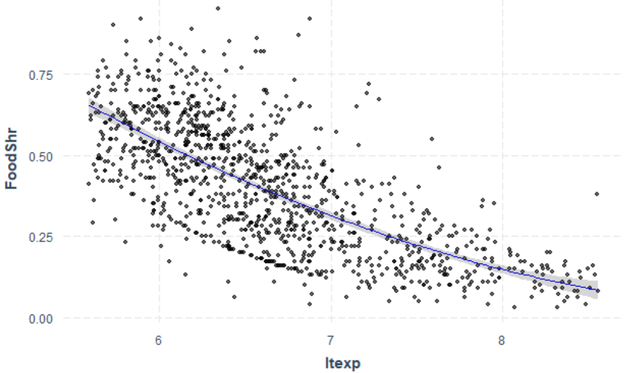


The fit (measured by R^2) is: $R^2 = 0.478$

Beyond linearity

- A Polynomial model may be better adapted: **Quadratic** model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

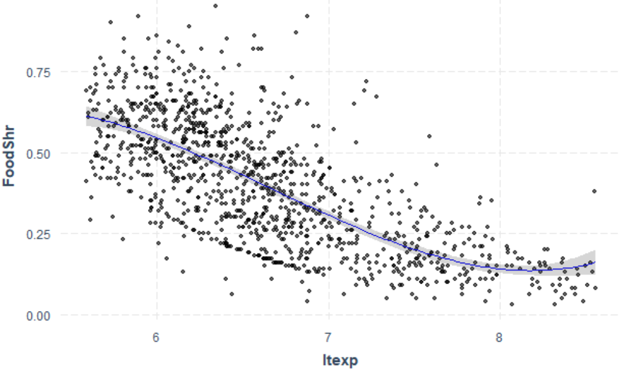


Do we have a better fit? $R^2 = 0.484$

BEYOND LINEARITY

- Polynomial may be better adapted: **Cubic** model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$



Do we have a better fit? R^2= 0.490

In practice

- Linear and polynomial models are determined by parameters $(\beta_0, \beta_2, \cdots, \beta_p)$

In practice

- Linear and polynomial models are determined by parameters $(\beta_0, \beta_2, \cdots, \beta_p)$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon$$

In practice

- Linear and polynomial models are determined by parameters $(\beta_0, \beta_2, \cdots, \beta_p)$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon$$

- How to choose the degree p ?

In practice

- Linear and polynomial models are determined by parameters $(\beta_0, \beta_2, \cdots, \beta_p)$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon$$

- How to choose the degree p ?
- Collinearity of x^p and x^q for $p \neq q$?

In practice

- Linear and polynomial models are determined by parameters $(\beta_0, \beta_2, \cdots, \beta_p)$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon$$

- How to choose the degree p ?
- Collinearity of x^p and x^q for $p \neq q$?

$\ldots$

In practice

- Linear and polynomial models are determined by parameters $(\beta_0, \beta_2, \cdots, \beta_p)$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon$$

- How to choose the degree p ?
- Collinearity of x^p and x^q for $p \neq q$?

  ...

$\hookrightarrow$ *How does that relates to the learning exercise?*

Nearest neighbors (k-NN)

- Other methods more flexible

Nearest neighbors (k-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)

Nearest neighbors (k-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)
  - $\hookrightarrow$ The goal is to estimate $f(\cdot)$ not β_s!

NEAREST NEIGHBORS (K-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)
  ↪ The goal is to estimate $f(\cdot)$ not β_s!
  Similar in spirit to "moving average" estimator

Nearest neighbors (k-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)
  - $\hookrightarrow$ The goal is to estimate $f(\cdot)$ not β_s!
    Similar in spirit to "moving average" estimator

$$\widehat{f}(x_i) = \frac{1}{k} \sum_{\substack{j \in \{k-\text{nearest} \\ \text{neighbours of } x_i\}}} y_j$$

Nearest neighbors (k-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)
  $\hookrightarrow$ The goal is to estimate $f(\cdot)$ not β_s!
  Similar in spirit to "moving average" estimator

$$\widehat{f}(x_i) = \frac{1}{k} \sum_{\substack{j \in \{k-nearest \\ neighbours\ of\ x_i\}}} y_j$$

k is the number of neighbors of x_i taken into account in the estimation.

Nearest neighbors (k-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)
  - $\hookrightarrow$ The goal is to estimate $f(\cdot)$ not β_s!
    Similar in spirit to "moving average" estimator

$$\widehat{f}(x_i) = \frac{1}{k} \sum_{\substack{j \in \{k-nearest \\ neighbours\ of\ x_i\}}} y_j$$

k is the number of neighbors of x_i taken into account in the estimation.

NEAREST NEIGHBORS (K-NN)

- Other methods more flexible
- Nearest neighbors (or k-NN)

  ↪ The goal is to estimate $f(\cdot)$ not β_s!
  Similar in spirit to "moving average" estimator

$$\widehat{f}(x_i) = \frac{1}{k} \sum_{\substack{j \in \{k-nearest \\ neighbours\ of\ x_i\}}} y_j$$

k is the number of neighbors of x_i taken into account in the estimation.

⊙ The method follows a very general idea:

Nearest neighbors (k-NN)

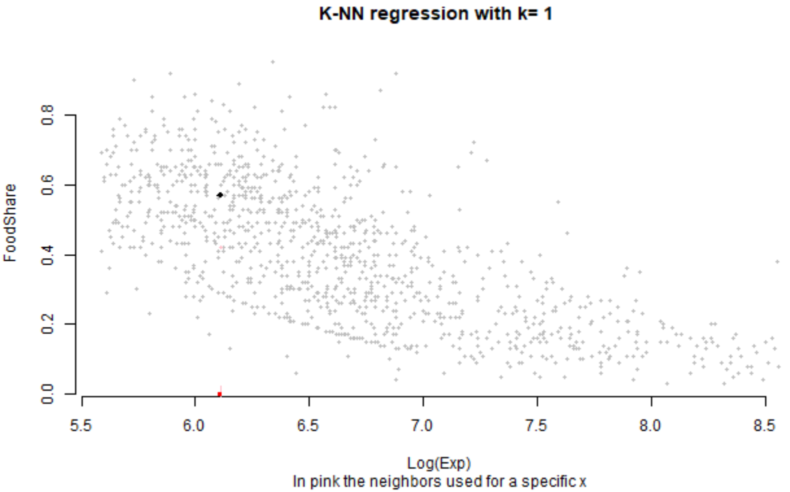
- Other methods more flexible
- Nearest neighbors (or k-NN)
  - ↪ The goal is to estimate $f(\cdot)$ not β_s!
    Similar in spirit to "moving average" estimator

$$\widehat{f}(x_i) = \frac{1}{k} \sum_{\substack{j \in \{k-nearest \\ neighbours\ of\ x_i\}}} y_j$$

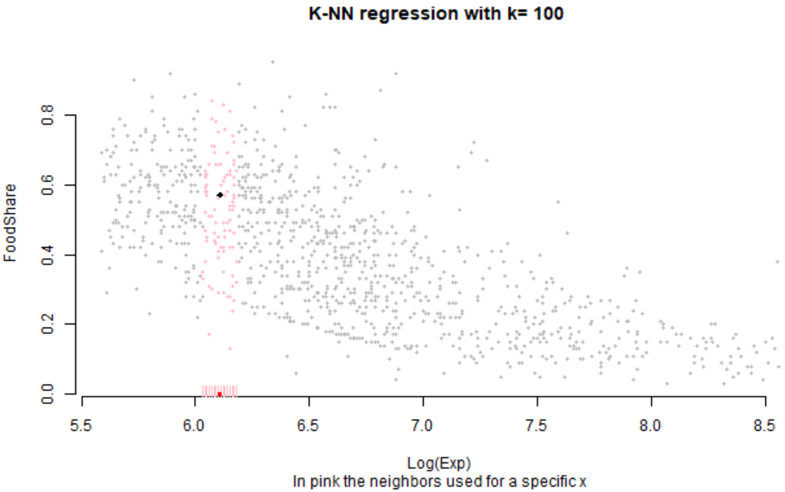
k is the number of neighbors of x_i taken into account in the estimation.

⊙ The method follows a very general idea:
"*Observations close in the x dimension should be close in the y dimension*"

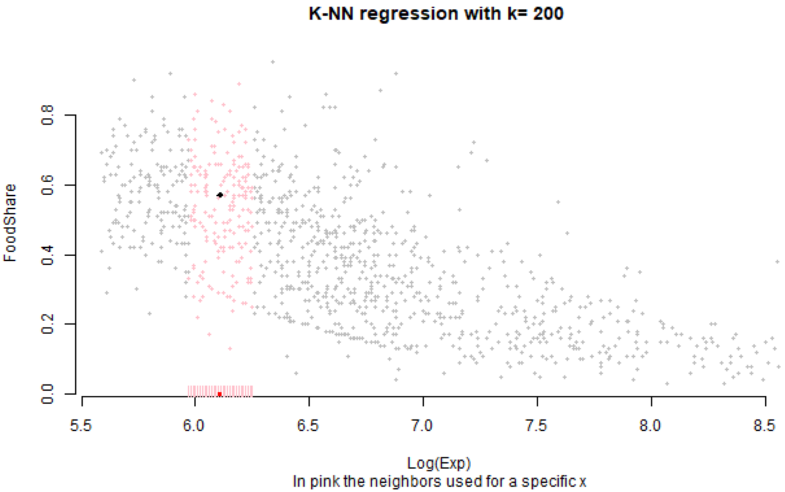
K-NN in practice



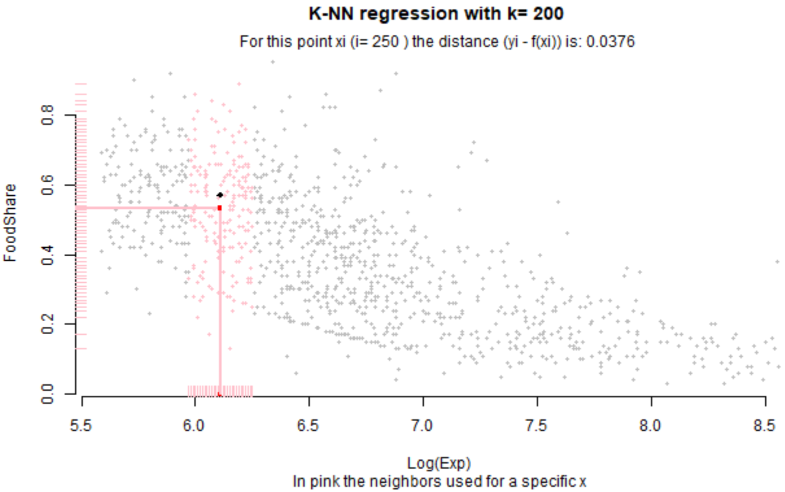
K-NN IN PRACTICE



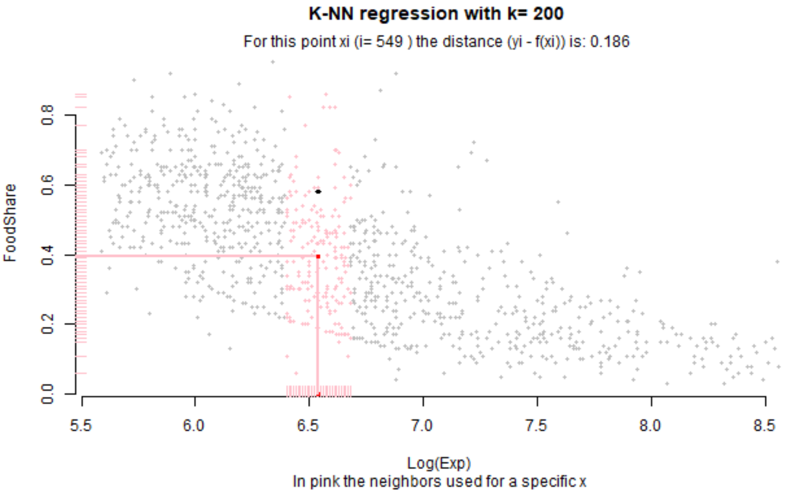
K-NN IN PRACTICE



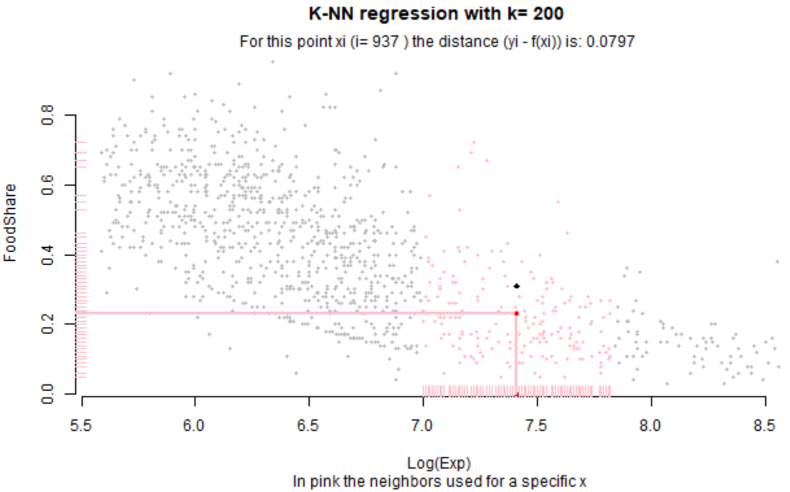
K-NN IN PRACTICE



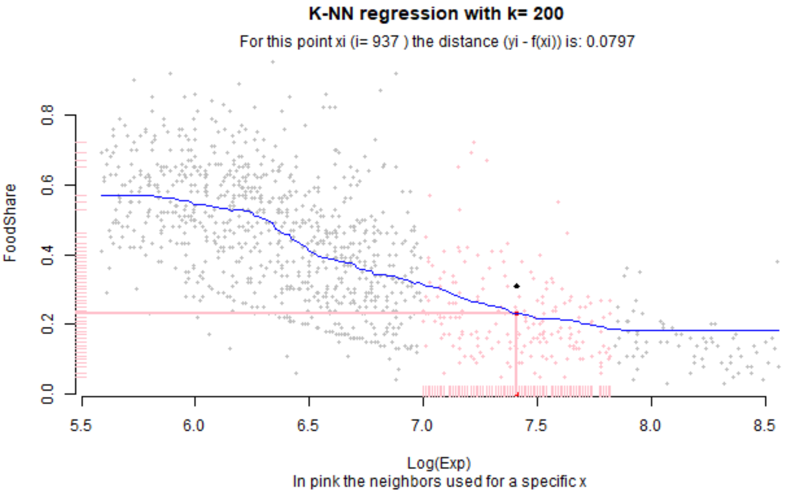
K-NN IN PRACTICE



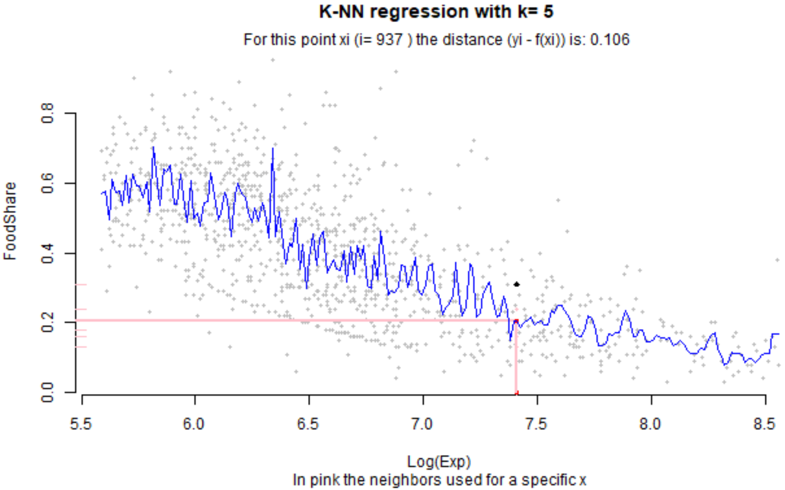
K-NN in practice



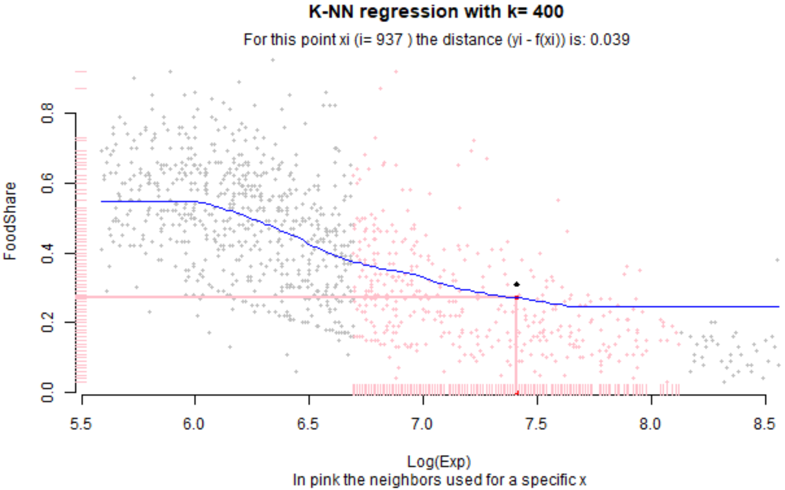
K-NN IN PRACTICE



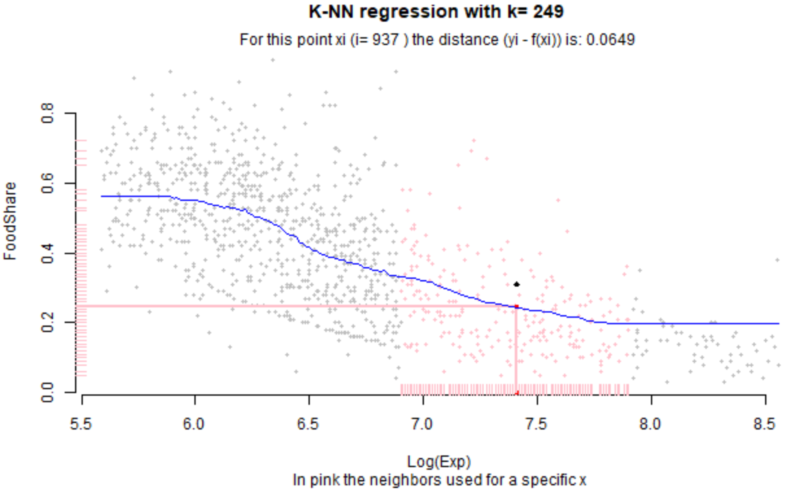
K-NN in practice: Choosing k



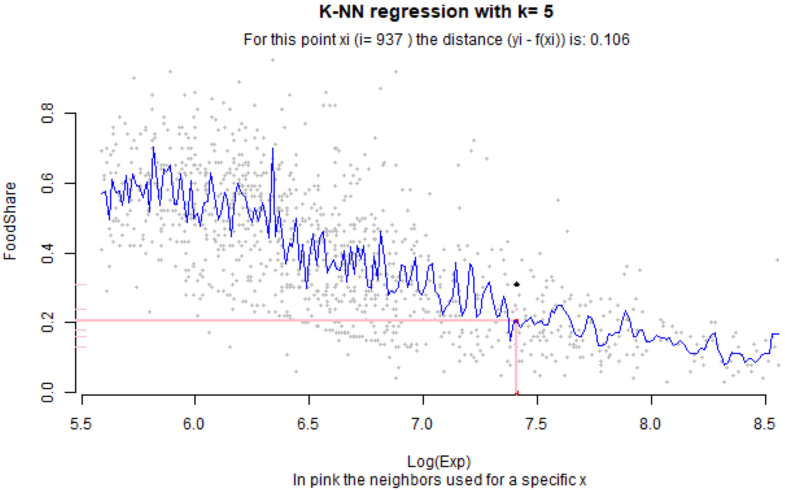
K-NN in practice: Choosing k



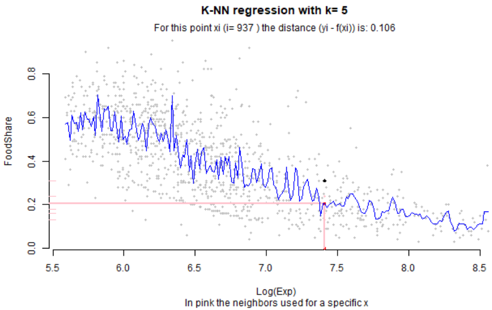
K-NN IN PRACTICE: CHOOSING K



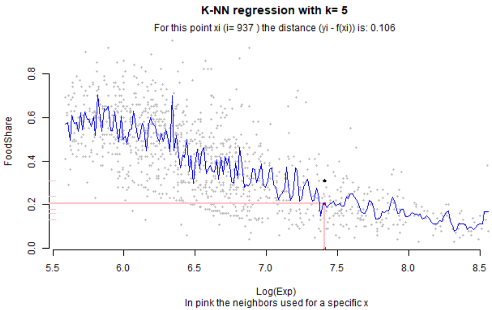
K-NN IN PRACTICE: CHOOSING K



K-NN in practice: Overfitting

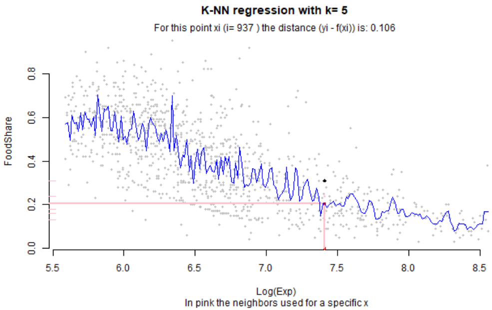


K-NN in practice: Overfitting



Overfitting has many consequences

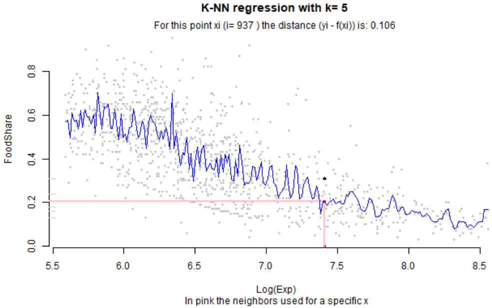
K-NN IN PRACTICE: OVERFITTING



Overfitting has many consequences

- The estimated curve follows the **data set** too closely

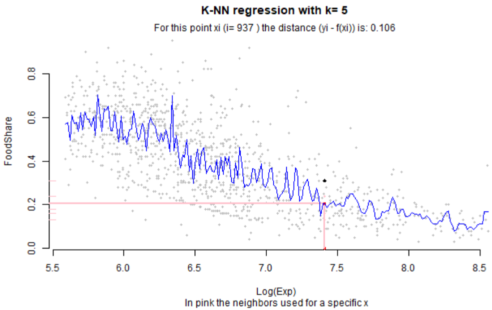
K-NN IN PRACTICE: OVERFITTING



Overfitting has many consequences

- The estimated curve follows the **data set** too closely
- The estimated curve follows the **errors** too closely

K-NN IN PRACTICE: OVERFITTING



Overfitting has many consequences

- The estimated curve follows the **data set** too closely
- The estimated curve follows the **errors** too closely
- The estimated function will not provide good estimates on **new observations**

STATISTICAL *vs* MACHINE LEARNING

What is the goal?

Statistical *vs* Machine Learning

What is the goal?

- If the goal is to formalize a model, one may focus on testing statistical properties, significance, relationships, ...

Statistical *vs* Machine Learning

What is the goal?

- If the goal is to formalize a model, one may focus on testing statistical properties, significance, relationships, ...

  $\hookrightarrow$ this is the purpose of **Statistical Learning**

STATISTICAL *vs* MACHINE LEARNING

What is the goal?

- If the goal is to formalize a model, one may focus on testing statistical properties, significance, relationships, ...

  $\hookrightarrow$ this is the purpose of **Statistical Learning**
- If the goal is to predict, one may focus on prediction accuracy

STATISTICAL *vs* MACHINE LEARNING

What is the goal?

- If the goal is to formalize a model, one may focus on testing statistical properties, significance, relationships, ...

  $\hookrightarrow$ this is the purpose of **Statistical Learning**
- If the goal is to predict, one may focus on prediction accuracy

  $\hookrightarrow$ this is the purpose of **Machine Learning**

STATISTICAL *vs* MACHINE LEARNING

What is the goal?

- If the goal is to formalize a model, one may focus on testing statistical properties, significance, relationships, ...

  $\hookrightarrow$ this is the purpose of **Statistical Learning**
- If the goal is to predict, one may focus on prediction accuracy

  $\hookrightarrow$ this is the purpose of **Machine Learning**
- Many statistical learning methods are relevant and useful to estimate $f(\cdot)$

Statistical *vs* Machine Learning

What is the goal?

- If the goal is to formalize a model, one may focus on testing statistical properties, significance, relationships, ...

  $\hookrightarrow$ this is the purpose of **Statistical Learning**
- If the goal is to predict, one may focus on prediction accuracy

  $\hookrightarrow$ this is the purpose of **Machine Learning**
- Many statistical learning methods are relevant and useful to estimate $f(\cdot)$
- In practice we'll use both tools to "*understand the data*"

What learning means?

The classical approach

- So far, we have estimated $f(\cdot)$ on **the whole data set**

What learning means?

The classical approach

- So far, we have estimated $f(\cdot)$ on **the whole data set**

What learning means?

The classical approach

- So far, we have estimated $f(\cdot)$ on **the whole data set**

- We have estimated $f(\cdot)$ by $\widehat{f}(\cdot)$ and minimized some cost function

What learning means?

The classical approach

- So far, we have estimated $f(\cdot)$ on **the whole data set**

- We have estimated $f(\cdot)$ by $\widehat{f}(\cdot)$ and minimized some cost function
- The data serve **both** for estimating $f(\cdot)$ **and** computing the prediction error

What learning means?

A different approach: *resampling*

- Our goal is evaluate the prediction accuracy of $\widehat{f}(\cdot)$ on a new, **unseen**, data set

What learning means?

A different approach: *resampling*

- Our goal is evaluate the prediction accuracy of $\widehat{f}(\cdot)$ on a new, **unseen**, data set
- Since we may not have **unseen** data, we will construct one

What learning means?

A different approach: *resampling*

- Our goal is evaluate the prediction accuracy of $\widehat{f}(\cdot)$ on a new, **unseen**, data set
- Since we may not have **unseen** data, we will construct one

What learning means?

A different approach: *resampling*

- Our goal is evaluate the prediction accuracy of $\widehat{f}(\cdot)$ on a new, **unseen**, data set
- Since we may not have **unseen** data, we will construct one

- Compare y_i with the prediction based on the validation set xs

Why different sets?

Estimating parameters using predictions accuracy

- When estimating $f(\cdot)$ on the whole data set, over-fitting may occur

Why different sets?

Estimating parameters using predictions accuracy

- When estimating $f(\cdot)$ on the whole data set, over-fitting may occur
- The validation set provides a good way to evaluate the prediction capabilities of a model and the prediction error on a new data set

Why different sets?

Estimating parameters using predictions accuracy

- When estimating $f(\cdot)$ on the whole data set, over-fitting may occur
- The validation set provides a good way to evaluate the prediction capabilities of a model and the prediction error on a new data set

Why different sets?

Estimating parameters using predictions accuracy

- When estimating $f(\cdot)$ on the whole data set, over-fitting may occur
- The validation set provides a good way to evaluate the prediction capabilities of a model and the prediction error on a new data set

- Prediction accuracy (using $\widehat{f}(\cdot)$) is then evaluated on the validation set **only**

Constructing training & validation sets

In practice, the validation set is not a block

CONSTRUCTING TRAINING & VALIDATION SETS

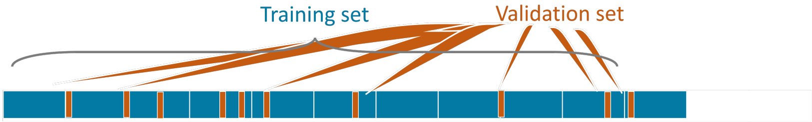
In practice, the validation set is not a block

- The validation set is constructed from a randomly drawn observations.

Constructing training & validation sets

In practice, the validation set is not a block

- The validation set is constructed from a randomly drawn observations.



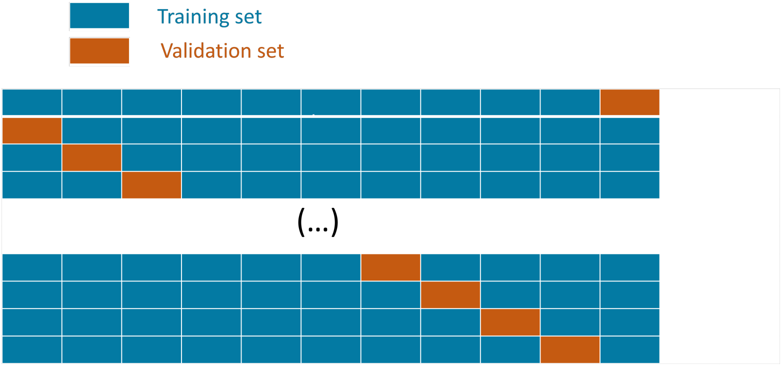
Many DIFFERENT SETS!

Using resampling methods to estimate the error on the prediction

Many DIFFERENT SETS!

Using resampling methods to estimate the error on the prediction

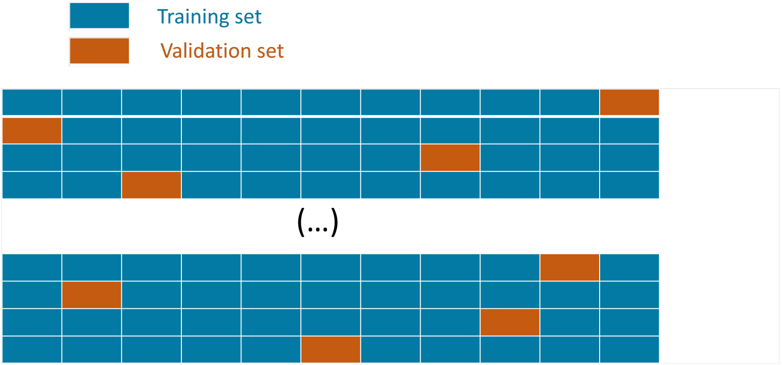
- Cross validation is used to select m-(training-validation) sets from the original data set (here again randomly)



Many DIFFERENT SETS!

Using resampling methods to estimate the error on the prediction

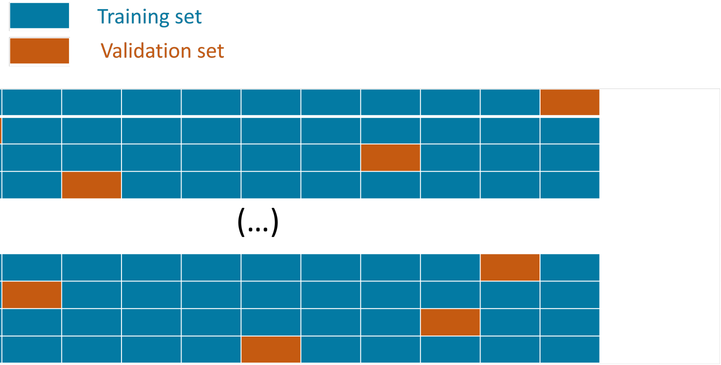
- Cross validation is used to select m-(training-validation) sets from the original data set (here again randomly)



Many DIFFERENT SETS!

Using resampling methods to estimate the error on the prediction

- Cross validation is used to select *m*-(training-validation) sets from the original data set (here again randomly)



Many DIFFERENT SETS!

m-fold Cross-Validation estimates the average prediction error on m different(training-validation) sets

Many DIFFERENT SETS!

m-fold Cross-Validation estimates the average prediction error on m different(training-validation) sets

- For each $(training - validation)$ set j, one can compute the MSE_j since the true ys are known on the validation set!

Many DIFFERENT SETS!

m-fold Cross-Validation estimates the average prediction error on m different(training-validation) sets

- For each $(training - validation)$ set j, one can compute the MSE_j since the true ys are known on the validation set!
- Cross Validation error is then:

$$CV_{(m)} = \frac{1}{m} \sum_{j=1}^{m} MSE_j$$

Many DIFFERENT SETS!

m-fold Cross-Validation estimates the average prediction error on m different(training-validation) sets

- For each $(training - validation)$ set j, one can compute the MSE_j since the true ys are known on the validation set!
- Cross Validation error is then:

$$CV_{(m)} = \frac{1}{m} \sum_{j=1}^{m} MSE_j$$

- $CV_{(m)}$ is a good estimate of the prediction error of the model

Many DIFFERENT SETS!

m-fold Cross-Validation estimates the average prediction error on m different(training-validation) sets

- For each $(training - validation)$ set j, one can compute the MSE_j since the true ys are known on the validation set!
- Cross Validation error is then:

$$CV_{(m)} = \frac{1}{m}\sum_{j=1}^{m} MSE_j$$

- $CV_{(m)}$ is a good estimate of the prediction error of the model
- $CV_{(m)}$ can serve to select and compare models

Many DIFFERENT SETS!

m-fold Cross-Validation estimates the average prediction error on m different(training-validation) sets

- For each $(training - validation)$ set j, one can compute the MSE_j since the true ys are known on the validation set!
- Cross Validation error is then:

$$CV_{(m)} = \frac{1}{m}\sum_{j=1}^{m} MSE_j$$

- $CV_{(m)}$ is a good estimate of the prediction error of the model
- $CV_{(m)}$ can serve to select and compare models

↪ Example: select k in k-NN regression

Tasks for Machine Learning

Machine Learning involves several tasks, some are time consuming

Tasks for Machine Learning

Machine Learning involves several tasks, some are time consuming

- Data collection (not treated here)

Tasks for Machine Learning

Machine Learning involves several tasks, some are time consuming

- Data collection (not treated here)
- Data organization (not treated here)

Tasks for Machine Learning

Machine Learning involves several tasks, some are time consuming

- Data collection (not treated here)
- Data organization (not treated here)
- Data cleaning (not treated here)

Tasks for Machine Learning

Machine Learning involves several tasks, some are time consuming

- Data collection (not treated here)
- Data organization (not treated here)
- Data cleaning (not treated here)
- Data visualization

Tasks for Machine Learning

Machine Learning involves several tasks, some are time consuming

- Data collection (not treated here)
- Data organization (not treated here)
- Data cleaning (not treated here)
- Data visualization
- Data analysis $\leftarrow$ this is the core of this course

Wrap-up

Wrap-up

- To *understand* the data, we use linear, polynomial, nonparametric models (k-NN) or other complex methods (including those for classification)

Wrap-up

- To *understand* the data, we use linear, polynomial, nonparametric models (k-NN) or other complex methods (including those for classification)
- More complex models $\nearrow$ accuracy, but $\rightarrow$ variability (variance) in the estimation

Wrap-up

- To *understand* the data, we use linear, polynomial, nonparametric models (k-NN) or other complex methods (including those for classification)
- More complex models $\nearrow$ accuracy, but $\rightarrow$ variability (variance) in the estimation
- There is a unavoidable **bias-variance** trade-off

Wrap-up

- To *understand* the data, we use linear, polynomial, nonparametric models (k-NN) or other complex methods (including those for classification)
- More complex models $\nearrow$ accuracy, but $\rightarrow$ variability (variance) in the estimation
- There is a unavoidable **bias-variance** trade-off
- Theory helps understanding but not in choosing the right model

Wrap-up

- To *understand* the data, we use linear, polynomial, nonparametric models (k-NN) or other complex methods (including those for classification)
- More complex models $\nearrow$ accuracy, but $\rightarrow$ variability (variance) in the estimation
- There is a unavoidable **bias-variance** trade-off
- Theory helps understanding but not in choosing the right model
- The (*train* + *validation*) sets approach is central in machine learning

Wrap-up

- To *understand* the data, we use linear, polynomial, nonparametric models (k-NN) or other complex methods (including those for classification)
- More complex models $\nearrow$ accuracy, but $\rightarrow$ variability (variance) in the estimation
- There is a unavoidable **bias-variance** trade-off
- Theory helps understanding but not in choosing the right model
- The (*train* + *validation*) sets approach is central in machine learning

$\hookrightarrow$ In a machine learning framework, the efficiency of the prediction will guide the choices, not the statistical properties!

[Q&A]

Write your questions in the chat

[Next week]

[Next week]

- Module 2: "Classification" (Examples of classifiers, Measures of fit, *Logit* as a classifier)

[Next week]

- Module 2: "Classification" (Examples of classifiers, Measures of fit, *Logit* as a classifier)
- Webinar on "Classification" Thursday, same time

[Next week]

- Module 2: "Classification" (Examples of classifiers, Measures of fit, *Logit* as a classifier)
- Webinar on "Classification" Thursday, same time
- Complete the activities before the webinar!

[Next week]

- Module 2: "Classification" (Examples of classifiers, Measures of fit, *Logit* as a classifier)
- Webinar on "Classification" Thursday, same time
- Complete the activities before the webinar!

Have a nice week!