

Machine Learning for Digital Soil Mapping-Part 2

Ruhollah Taghizadeh
Soil Science and Geomorphology

E-mail: ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de

Content

☐ Part 1

- Digital Soil Mapping
- Machine Learning

☐ Part 2

- Lets practice!

SCORPAN model

$$S = f(s, c, o, r, p, a, n) + \varepsilon$$

S : Soil, at a specific point in space and time: soil classes, **Sc** or soil attributes, **Sa**

From Jenny's Equation

c : climate, climate properties of the environment;

o : organisms, vegetation;

r : topography, landscape attributes;

p : parent material, lithology;

a : age or time factor;

Additions:

s : soil, prior knowledge of the soil at a point;

n : space, relative spatial position;

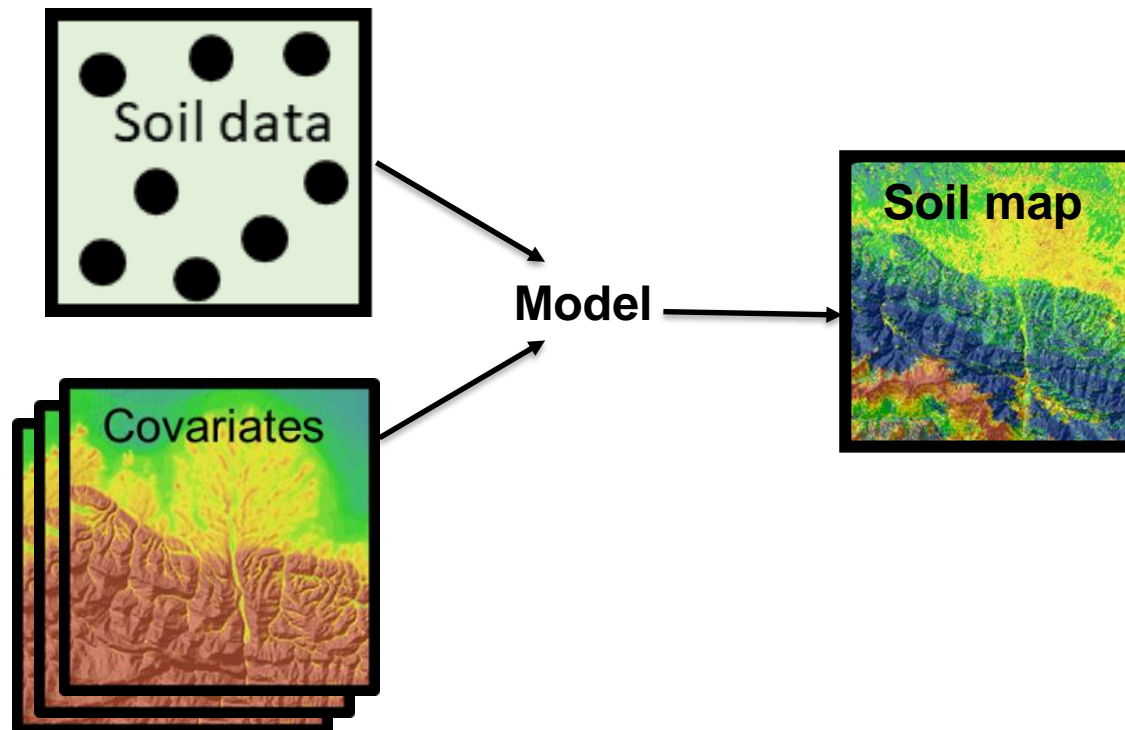
ε : auto-correlated random spatial variation.

f() : Quantitative function **f** linking **S** to **scorpan** factors

SCORPAN model

$$\boxed{S} = \boxed{f}(\boxed{s, c, o, r, p, a, n}) + \varepsilon$$

Soil data Model Covariates ε spatially dependent residuals



SCORPAN Model

$$S = \textcircled{f}(s, c, o, r, p, a, n) + \varepsilon$$



Machine learning

$f()$: Quantitative function f linking S to $scorpan$ factors

Sequence of DSM Steps



Select Covariates

Sampling Locations

Data Preparation

Model Training and Validation

Predictions



1

Environmental covariates, relevant as predictors of soil property/class, are derived from remote sensing, digital elevation, climatic datasets, ...

2

Soil samples are collected at the specified locations (e.g., Latin hypercube sampling) and soil property is measured in the laboratory.

3

Intersecting the covariates with the soil point observations.

4

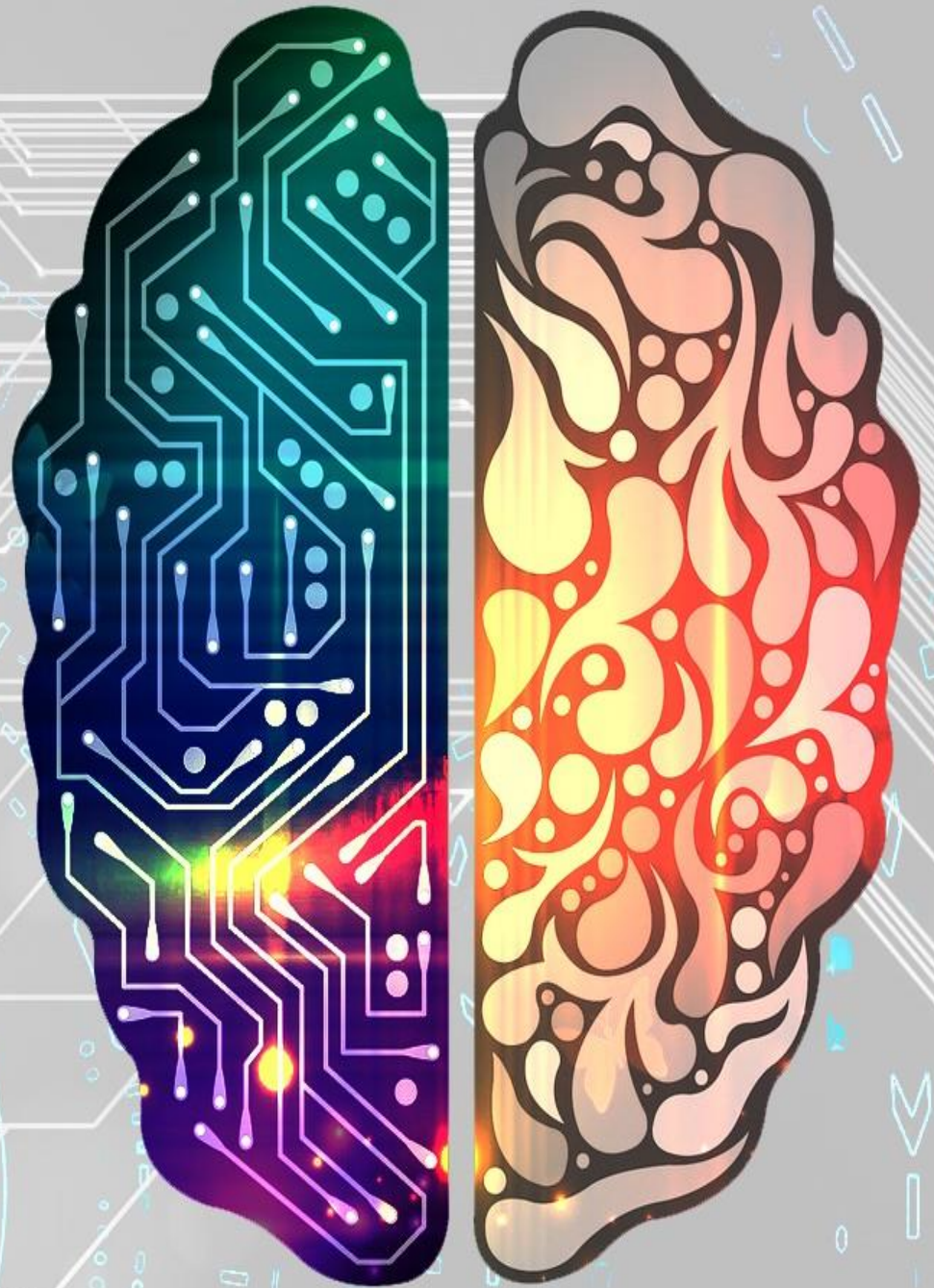
Machine learning models (e.g., random forest) are trained using training data, and accuracy assessment is carried out using the test data set.

5

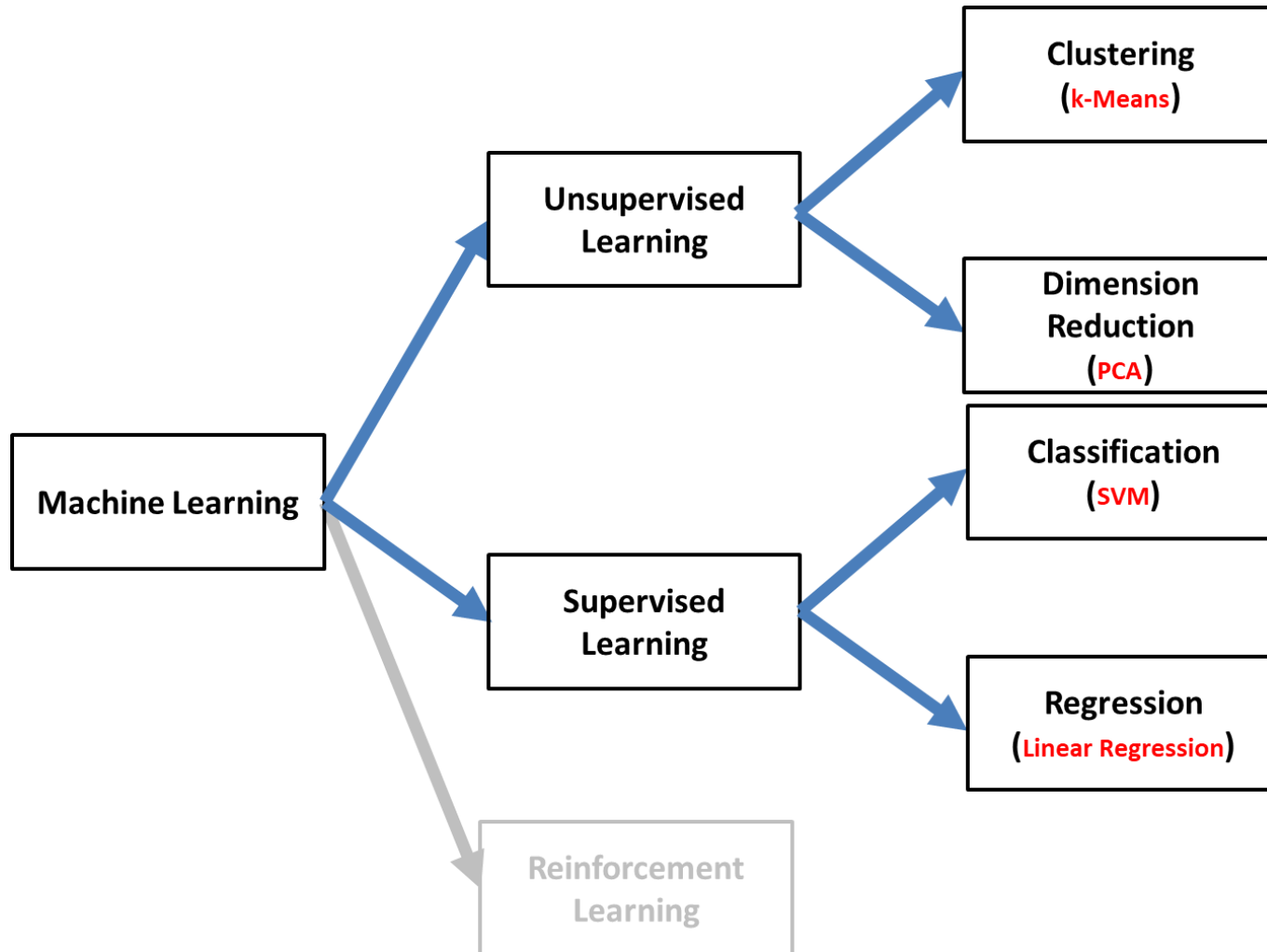
The ML models are applied to the entire study area in order to produce a soil property/class map.

Introduction: some terms

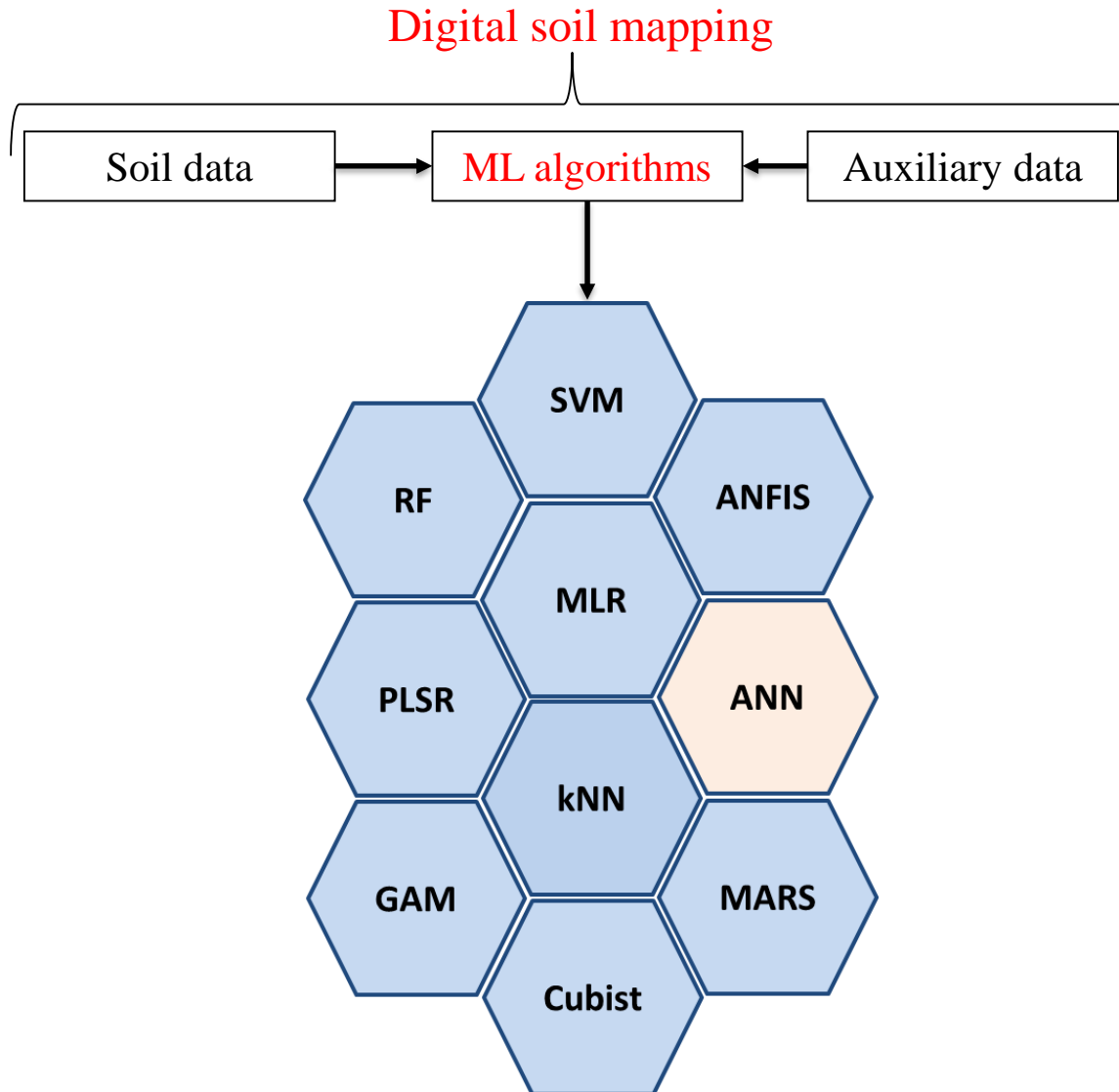
MACHINE LEARNING



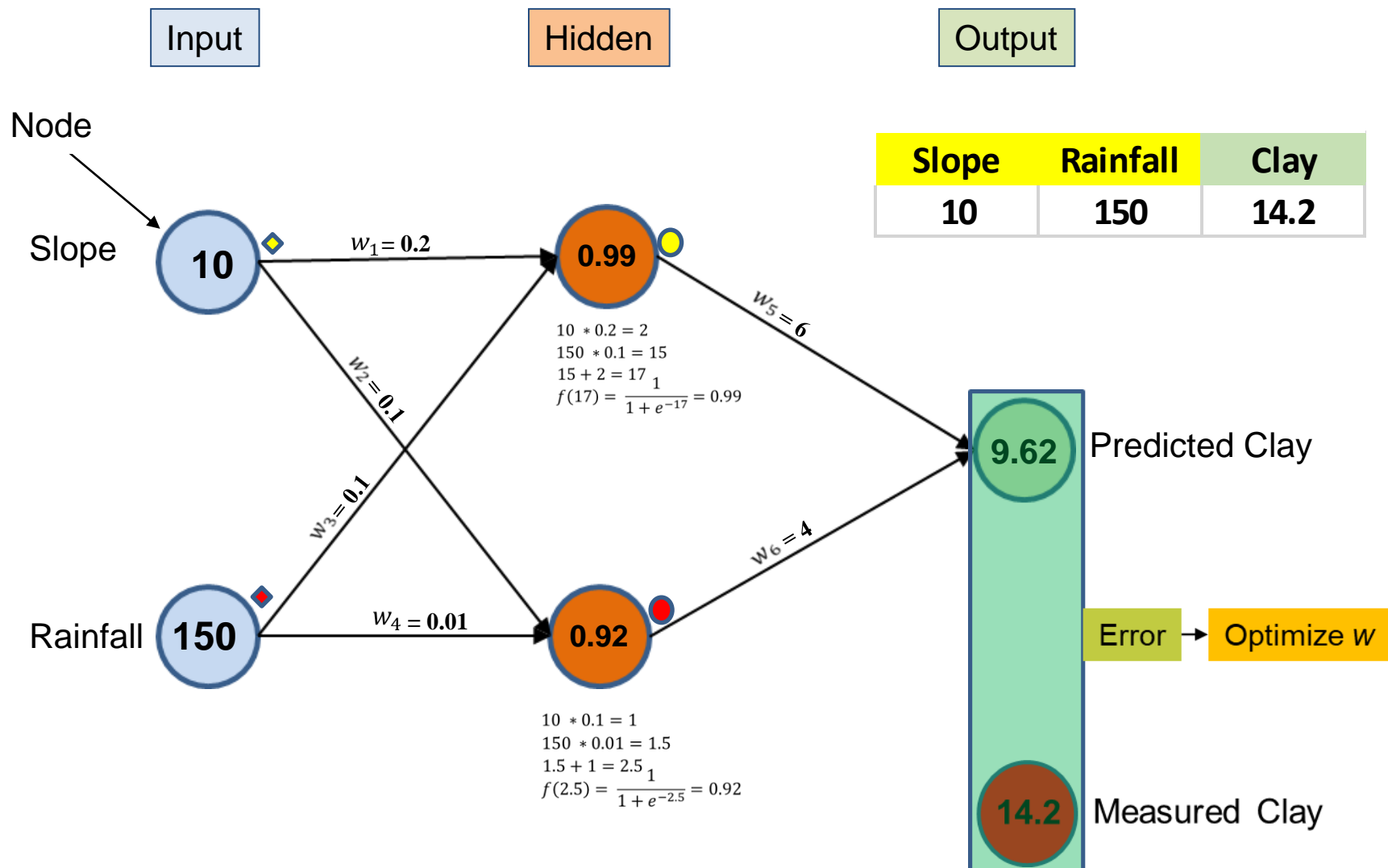
Types of Machine Learning



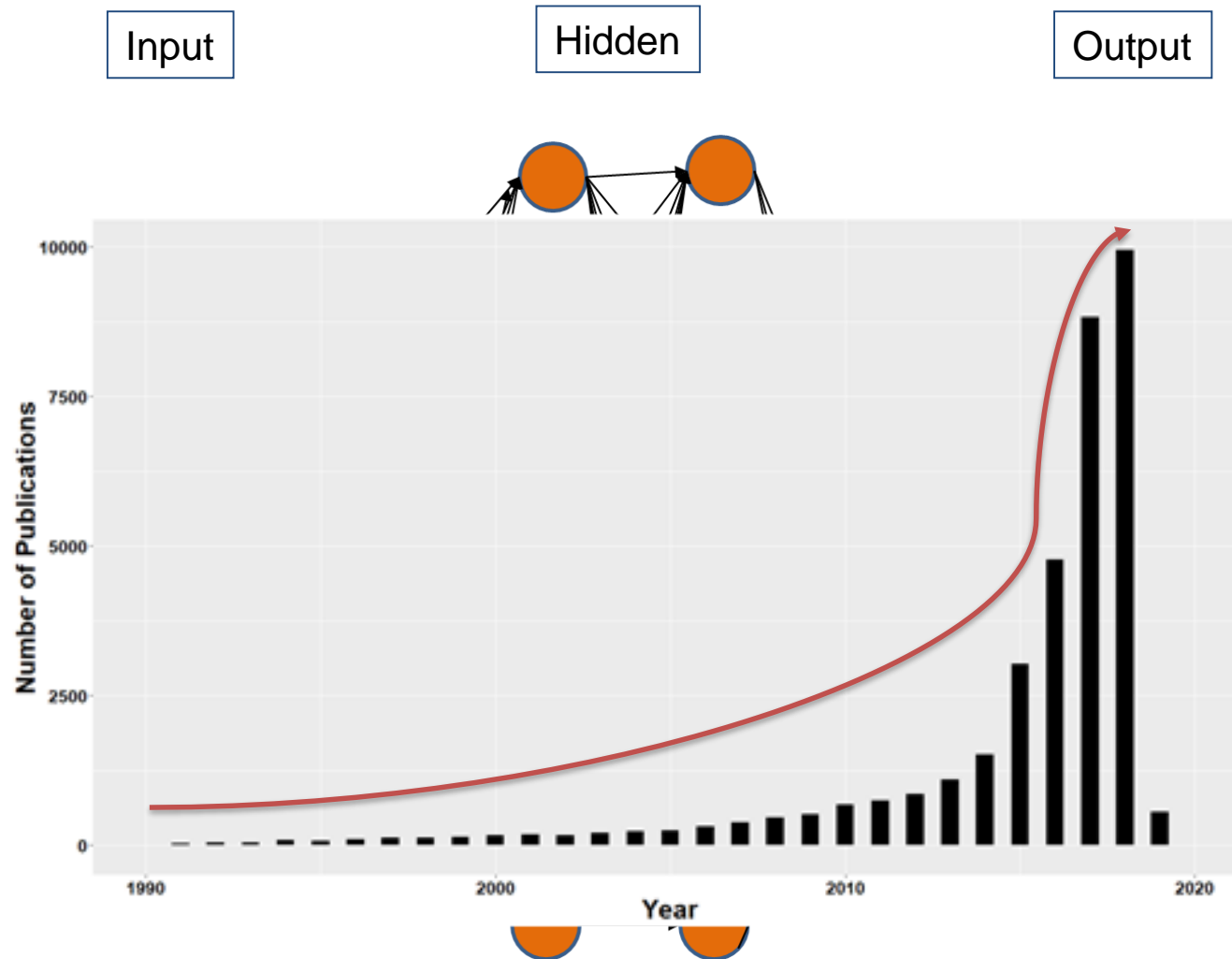
□ Different ML algorithms in DSM



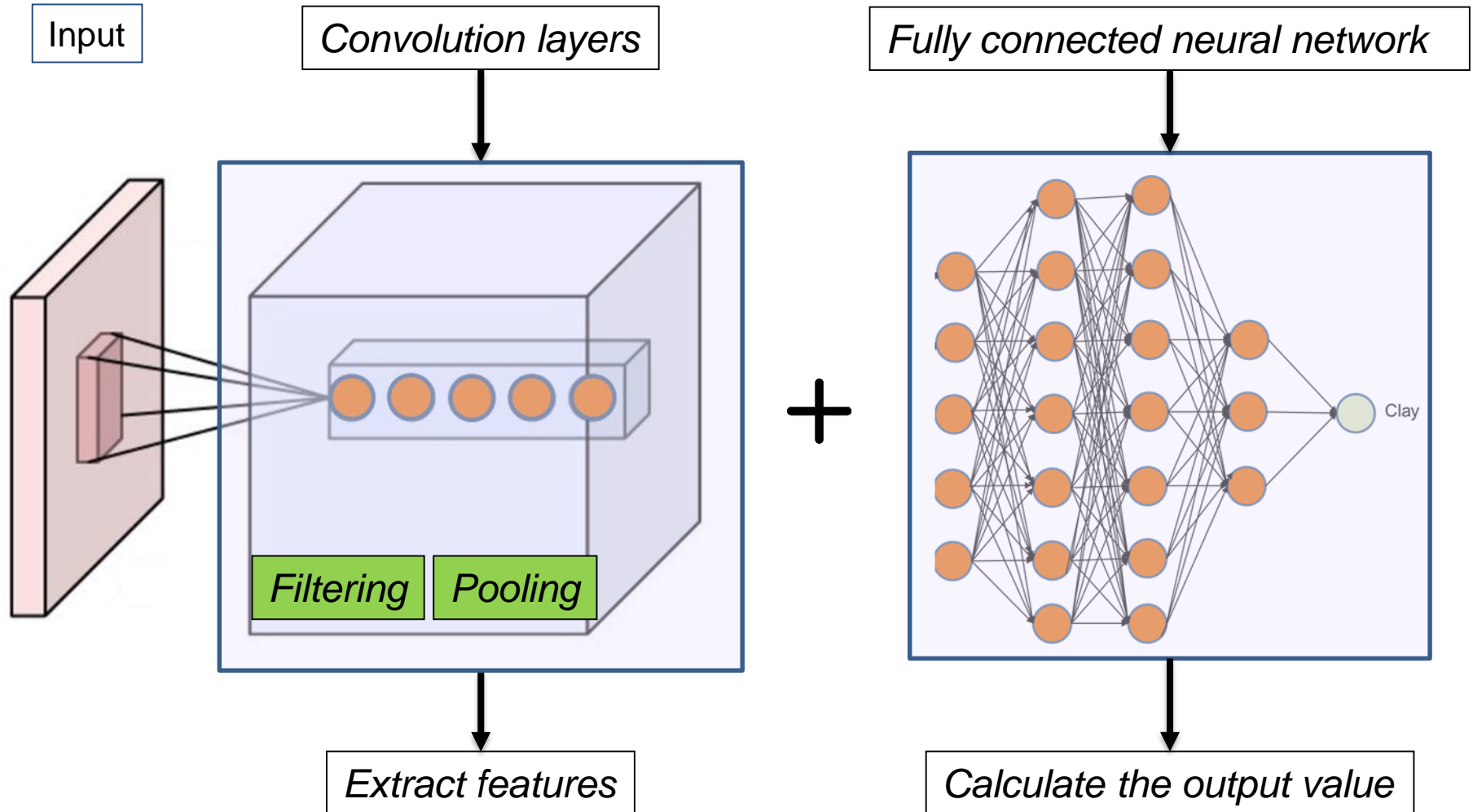
□ Artificial neural networks (ANN)



□ Deep artificial neural networks (DL)

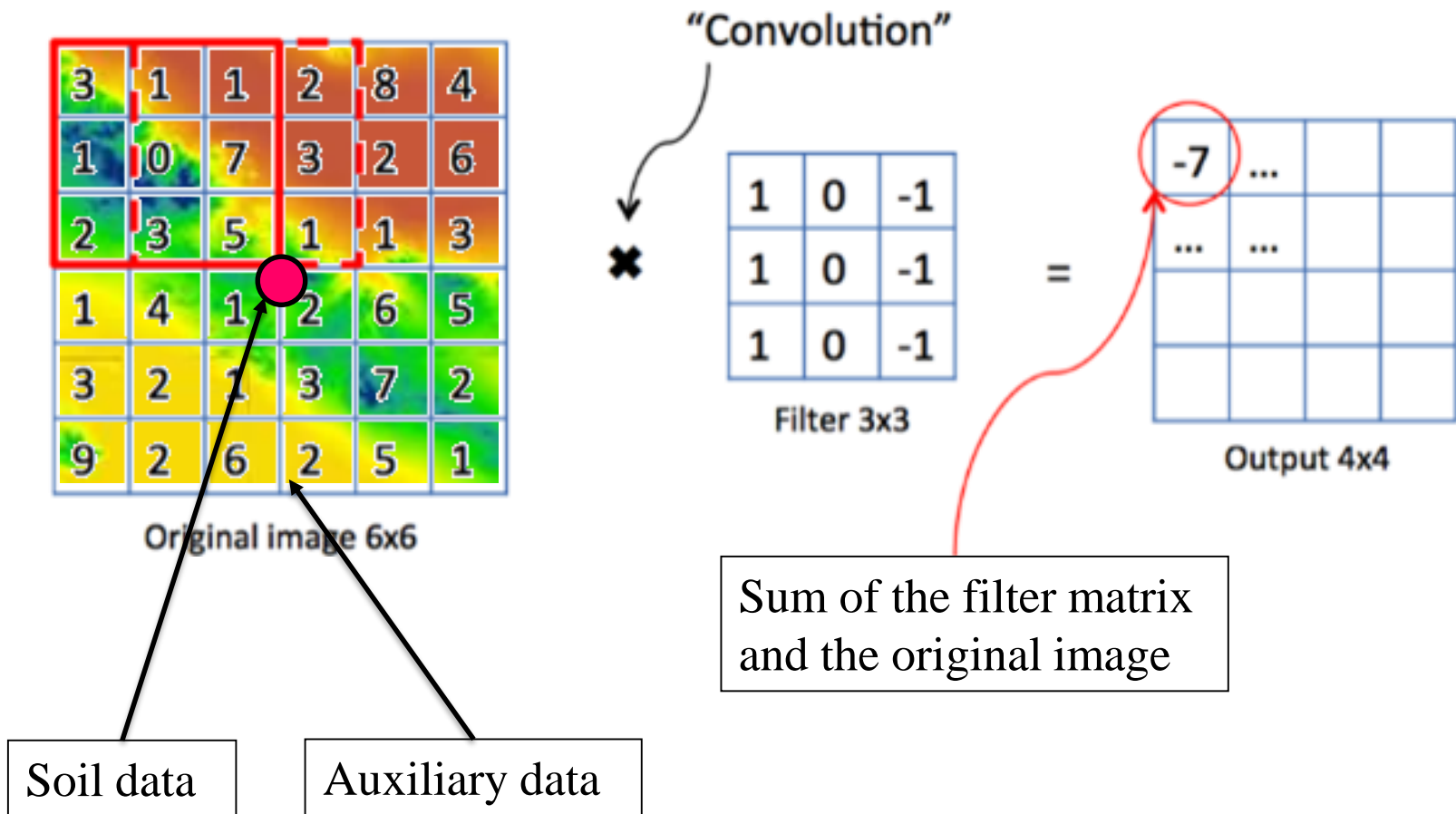


□ Convolutional neural networks (CNN)



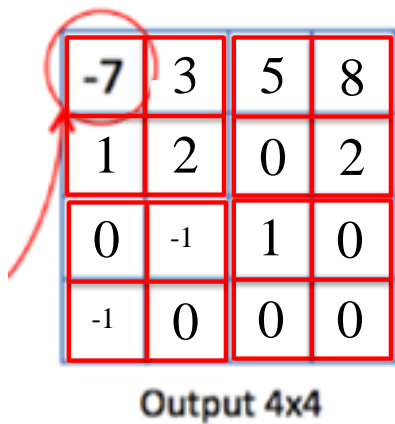
□ Convolutional neural networks (CNN)

- Convolutional layers: Filters

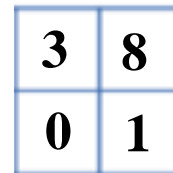


□ Convolutional neural networks (CNN)

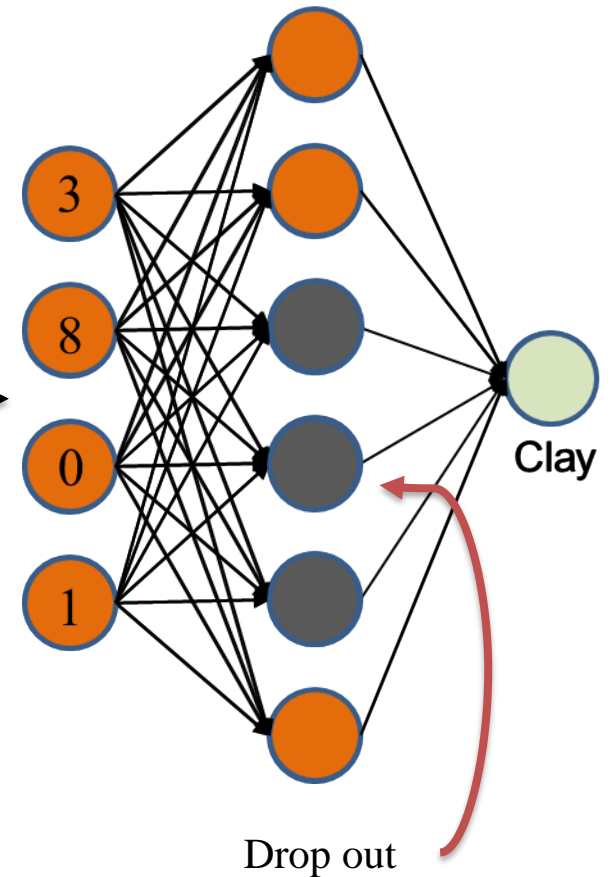
- Convolutional layers: Pooling



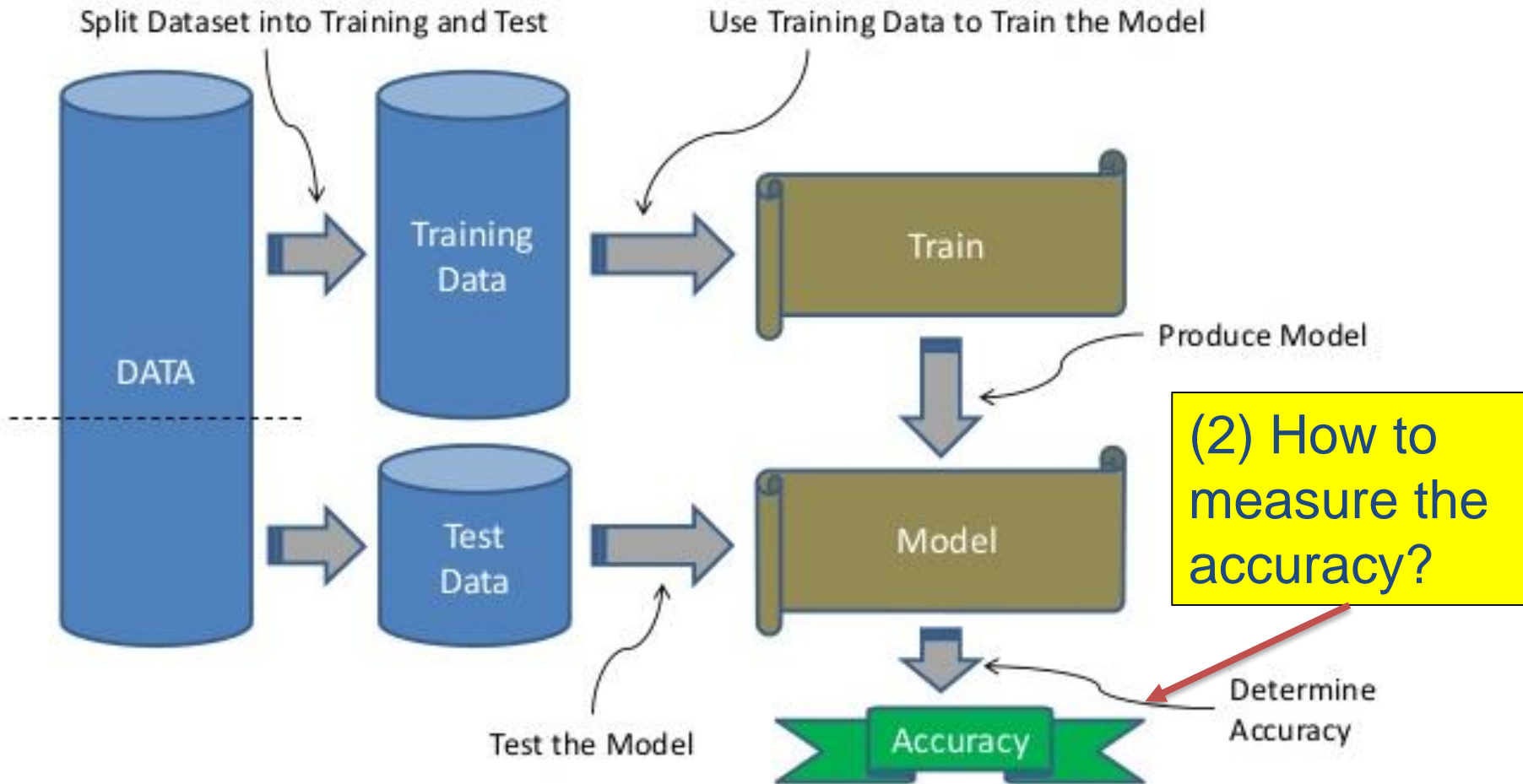
Max Pooling



- Fully connected neural network



Accuracy Assessment of Models



How to measure the accuracy?

- Performance Metrics:

Regression	Classification
<ul style="list-style-type: none">• Mean Absolute Error (MAE)• Root Mean Squared Error (RMSE)• R-Squared and Adjusted R-Squared	<ul style="list-style-type: none">• Recall• Precision• F1-Score• Accuracy• Area Under the Curve (AUC)

Performance Metrics for Classification

- **Confusion Matrix:** is one of the most intuitive and easiest metrics used for finding the correctness and accuracy of the model.

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	TP correct	FP
	NEGATIVE	FN	TN correct

where TP, TN, FP and FN are true positive, true negative, false positive and false negative, respectively

Performance Metrics for Classification

- Confusion Matrix:

● Class A: Aridisols

● Class B: Entisols

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols		
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols

 Class B: Entisols

Actual Test data

Predicted Test data

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols		
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols

 Class B: Entisols

Actual Test data

1 2 3 4 5

Predicted Test data

1 2 3 4 5

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols	1	
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols

 Class B: Entisols

Actual Test data

    
2

Predicted Test data

    
2

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols	1 1	
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols

 Class B: Entisols

Actual Test data

Predicted Test data

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols		
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols

 Class B: Entisols

Actual Test data

1 2 3 4 5

Predicted Test data

1 2 3 4 5

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols		
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols






 Class B: Entisols

Actual Test data

Predicted Test data

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols	 	
Class B: Entisols		

Performance Metrics for Classification

- Confusion Matrix:

 Class A: Aridisols





 Class B: Entisols

Actual Test data

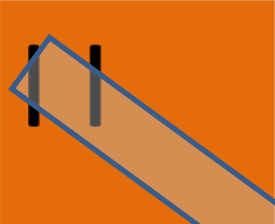


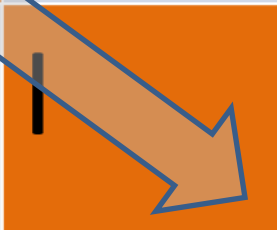
Predicted Test data

	Class A: Aridisols	Class B: Entisols
Class A: Aridisols		
Class B: Entisols		

Performance Metrics for Classification

- **Overall Accuracy:** is a metric calculating the classifier overall accuracy





	Class A: Aridisoils	Class B: Entisols
Class A: Aridisoils		
Class B: Entisols		

$$OA = \left(\frac{\text{Correctly classified}}{\text{Total number of test data}} \right) * 100$$

$$OA = \left(\frac{2 + 1}{5} \right) * 100 = 60\%$$

Performance Metrics for Classification

- **Precision:** is the proportion of those predicted instances that are correctly classified

	Class A: Aridisoils	Class B: Entisols
Class A: Aridisoils		
Class B: Entisols		

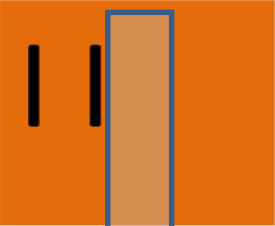
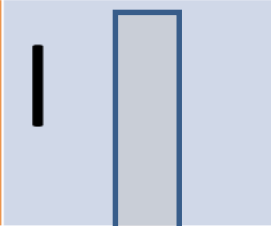
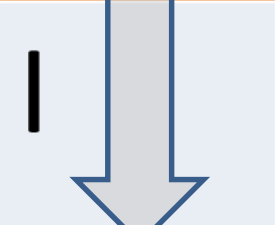
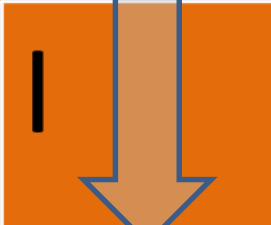
$$Pr = \left(\frac{\text{Correctly classified for each class}}{\text{Total number of row data}} \right) * 100$$

$$Pr(class A) = \left(\frac{2}{2 + 1} \right) * 100 = 66\%$$

$$Pr(class B) = \left(\frac{1}{1 + 1} \right) * 100 = 50\%$$

Performance Metrics for Classification

- Recall:** is the proportion of those instances that are correctly classified

	Class A: Aridisoils	Class B: Entisoils
Class A: Aridisoils		
Class B: Entisoils		

$$Re = \left(\frac{\text{Correctly classified for each class}}{\text{Total number of coloum data}} \right) * 100$$

$$Re(class A) = \left(\frac{2}{2 + 1} \right) * 100 = 66\%$$

$$Re(class B) = \left(\frac{1}{1 + 1} \right) * 100 = 50\%$$

Performance Metrics for Classification

- **F-score:** the F-score is the harmonic mean of precision and recall

$$F - score = \left(\frac{2 \times Precision \times Recall}{Precision + Recall} \right) * 100$$

Metric	Formula
Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
Error rate	$ERR = \frac{FP+FN}{TP+TN+FP+FN}$
Precision	$PRC = \frac{TP}{TP+FP}$
Sensitivity	$SNS = \frac{TP}{TP+FN}$
Specificity	$SPC = \frac{TN}{TN+FP}$
ROC	$ROC = \frac{\sqrt{SNS^2+SPC^2}}{\sqrt{2}}$
F_1 score	$F_1 = 2 \frac{PRC \cdot SNS}{PRC + SNS}$
Geometric Mean	$GM = \sqrt{SNS \cdot SPC}$

Practice

Import covariates

Import point data

Overlay point data on covariates

Make a geodatabase

Split the geodatabase to training and testing sets

Train machine learning models

Test machine learning models

Predict soil maps



1. Open Rstudio

2. Clear Project List

The screenshot displays the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar below the menu bar contains icons for creating a new file, opening a file, saving, and navigating. The main editor pane on the left shows a script titled 'Untitled1*' with the following R code:

```
1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5
```

The console pane at the bottom left shows the R startup message:

```
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

The right-hand pane shows the 'Environment' tab, which currently displays 'Global Environment'. A red arrow points from the '2. Clear Project List' header to the 'Clear Project List' option in the 'Project' menu, which is open. The menu options are:

- New Project...
- Open Project...
- Open Project in New Session...
- Close Project
- Clear Project List
- Project Options...

The Windows taskbar at the bottom shows the Start button, a search bar, and several open applications including File Explorer, Google Chrome, and RStudio. The system clock indicates 4:35 PM on 2/9/2021.

3. New Project

The screenshot displays the RStudio application window. The top menu bar includes File, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu bar is a toolbar with icons for creating a new file, opening a file, saving, and running code. The main editor window shows a script titled 'Untitled1*' with the following R code:

```
1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5
```

The console window at the bottom left shows the R startup message:

```
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

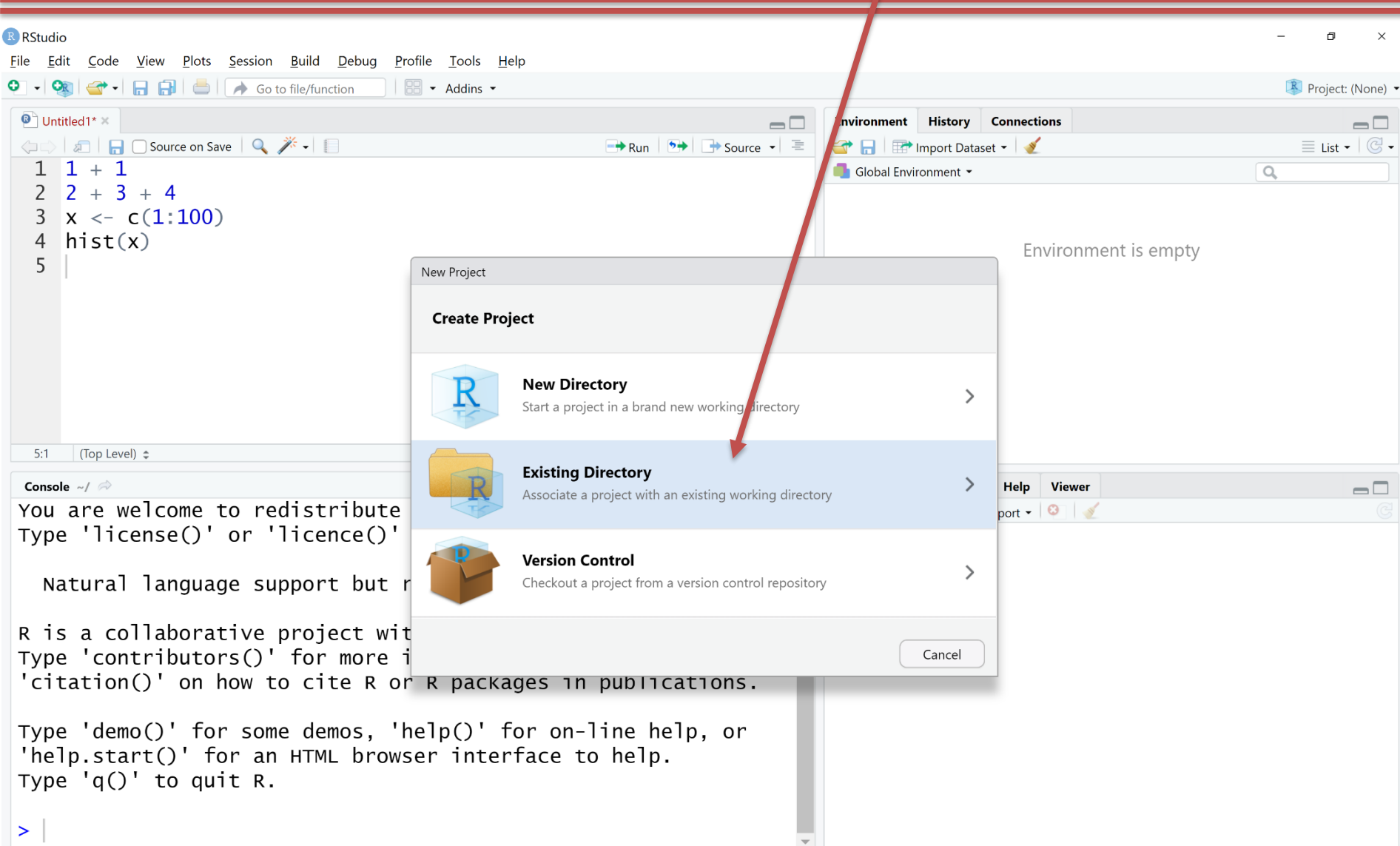
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

The right-hand pane shows the 'Environment' tab, which displays 'Global Environment'. A red arrow points to the 'New Project...' button in the top right corner of the RStudio window. The 'Project: (None)' dropdown menu is open, showing options: New Project..., Open Project..., Open Project in New Session..., Close Project, Clear Project List (highlighted), and Project Options...

The bottom status bar shows the Windows taskbar with the search bar and various application icons, including RStudio and the Windows Explorer. The system clock indicates 4:35 PM on 2/9/2021.

4. Existing Directory



The image shows the RStudio interface with a 'New Project' dialog box open. A red arrow points to the 'Existing Directory' option. The dialog box has three options: 'New Directory', 'Existing Directory', and 'Version Control'. The 'Existing Directory' option is highlighted. The background shows the RStudio editor with a script file 'Untitled1*' containing R code, the console with the R welcome message, and the environment pane showing 'Global Environment'.

RStudio Interface:

- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help
- Toolbar:** Go to file/function, Addins
- Environment Pane:** Project: (None), Global Environment
- Console:** You are welcome to redistribute... Type 'license()' or 'licence()'... Natural language support but r... R is a collaborative project with... Type 'contributors()' for more i... 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

New Project Dialog Box:

- Create Project**
- New Directory:** Start a project in a brand new working directory
- Existing Directory:** Associate a project with an existing working directory (highlighted)
- Version Control:** Checkout a project from a version control repository
- Cancel**

5. Find the folder and find **Day_05_Machine Learning-P2** and open

The screenshot shows the RStudio interface with two dialog boxes open. The 'Choose Directory' dialog box is in the foreground, showing a file explorer view of the 'Soil and Water Research' directory. The '03_Day_Covariates' folder is selected. The 'Create Project' dialog box is in the background, showing the 'From Existing Directory' tab. The 'Browse...' button is highlighted. Red arrows point from the text 'Find the folder' to the '03_Day_Covariates' folder and from the text 'find Day_05_Machine Learning-P2' to the 'Browse...' button.

Choose Directory

Desk... > Soil and Water Re...

Search: Soil and Water Resear...

Organize · New folder

This PC

- 3D Objects
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- OS (C:)

Name	Date modified	Type
03_Day_Covariates	2/9/2021 4:14 PM	File folder
CODES	2/7/2021 8:35 AM	File folder
Literature	2/6/2021 11:03 PM	File folder
Presentations	2/6/2021 11:03 PM	File folder
software	2/6/2021 9:52 PM	File folder

Folder: 03_Day_Covariates

Open Cancel

From Existing Directory

Directory:

Browse...

Environment

History

Connections

Project: (None)

Import Dataset

Global Environment

Environment is empty

Help

Viewer

port

Create Project Cancel

Type 'license()' or 'licence()' for license information. See <https://r-project.org/licenses/2018-environment-licenses/> for details.

Natural language support but running in an English locale.

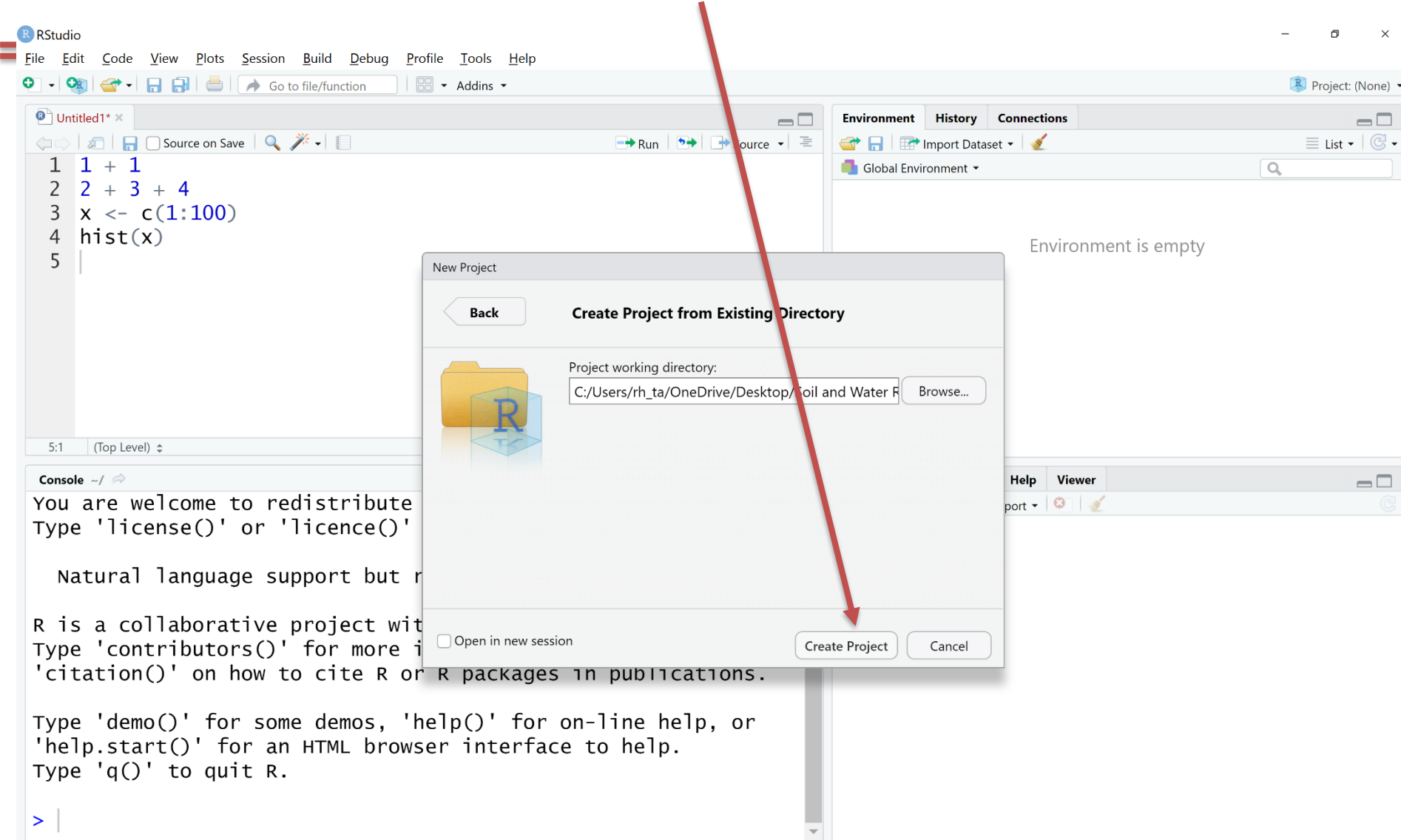
R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

> |

6. Create project

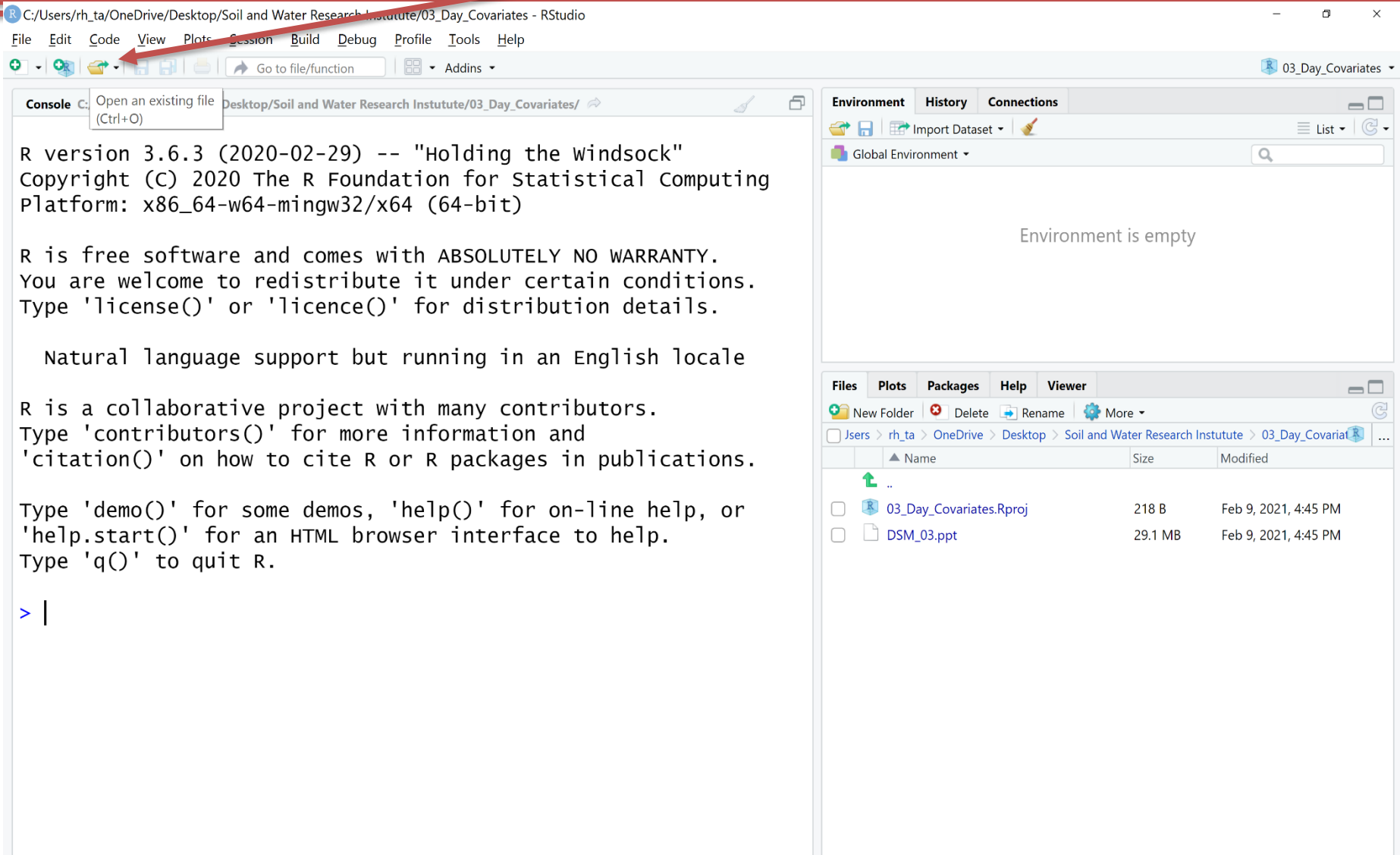


The screenshot shows the RStudio interface with the 'New Project' dialog box open. A red arrow points from the title '6. Create project' to the 'Create Project' button in the dialog. The dialog is titled 'New Project' and has a 'Back' button. It contains a folder icon with the R logo and the text 'Create Project from Existing Directory'. Below this, it says 'Project working directory:' followed by the path 'C:/Users/rh_ta/OneDrive/Desktop/Soil and Water R' and a 'Browse...' button. At the bottom, there is a checkbox labeled 'Open in new session' and two buttons: 'Create Project' and 'Cancel'. The background shows the RStudio editor with a script file 'Untitled1*' containing R code, the Environment pane showing 'Global Environment', and the Console pane with the R startup message.

```
1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5
```

Console ~/
You are welcome to redistribute
Type 'license()' or 'licence()' for
Natural language support but r
R is a collaborative project with
Type 'contributors()' for more i
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>

7. Open an existing file and find **05_R_Cov.R**



The screenshot shows the RStudio interface. The title bar indicates the current project is located at `C:/Users/rh_ta/OneDrive/Desktop/Soil and Water Research Institute/03_Day_Covariates - RStudio`. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for opening files, saving, and other standard functions. A red arrow points from the title bar area to the 'Open an existing file (Ctrl+O)' button in the toolbar.

The Console window displays the following text:

```
R version 3.6.3 (2020-02-29) -- "Holding the windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

The Environment window shows 'Global Environment' and 'Environment is empty'. The Files window shows the project structure:

Name	Size	Modified
..		
03_Day_Covariates.Rproj	218 B	Feb 9, 2021, 4:45 PM
DSM_03.ppt	29.1 MB	Feb 9, 2021, 4:45 PM

8. Open 05_R_Cov.R

