

材料信息学计算机课堂 实验

2019-6-12

Classroom lab I: regression

- 利用Microsoft Azure云计算平台完成数据的回归分析
- Sign in Azure Machine Learning Studio:
<https://studio.azureml.net>
- 根据回归分析结果分析各参数的相对重要性
- Example data (data for exercise.xlsx): download from <http://cms.sjtu.edu.cn/download.html>
- For tutorial: <https://docs.microsoft.com/en-us/azure/machine-learning/studio/create-experiment>

Import data

 Automobile price data (Raw)

Prepare the data


 Select Columns in Dataset
exclude normalized losses

 Clean Missing Data

Define the features


 Select Columns in Dataset

 Linear Regression


 Split Data
85% for training (left port)
15% for scoring (right port)

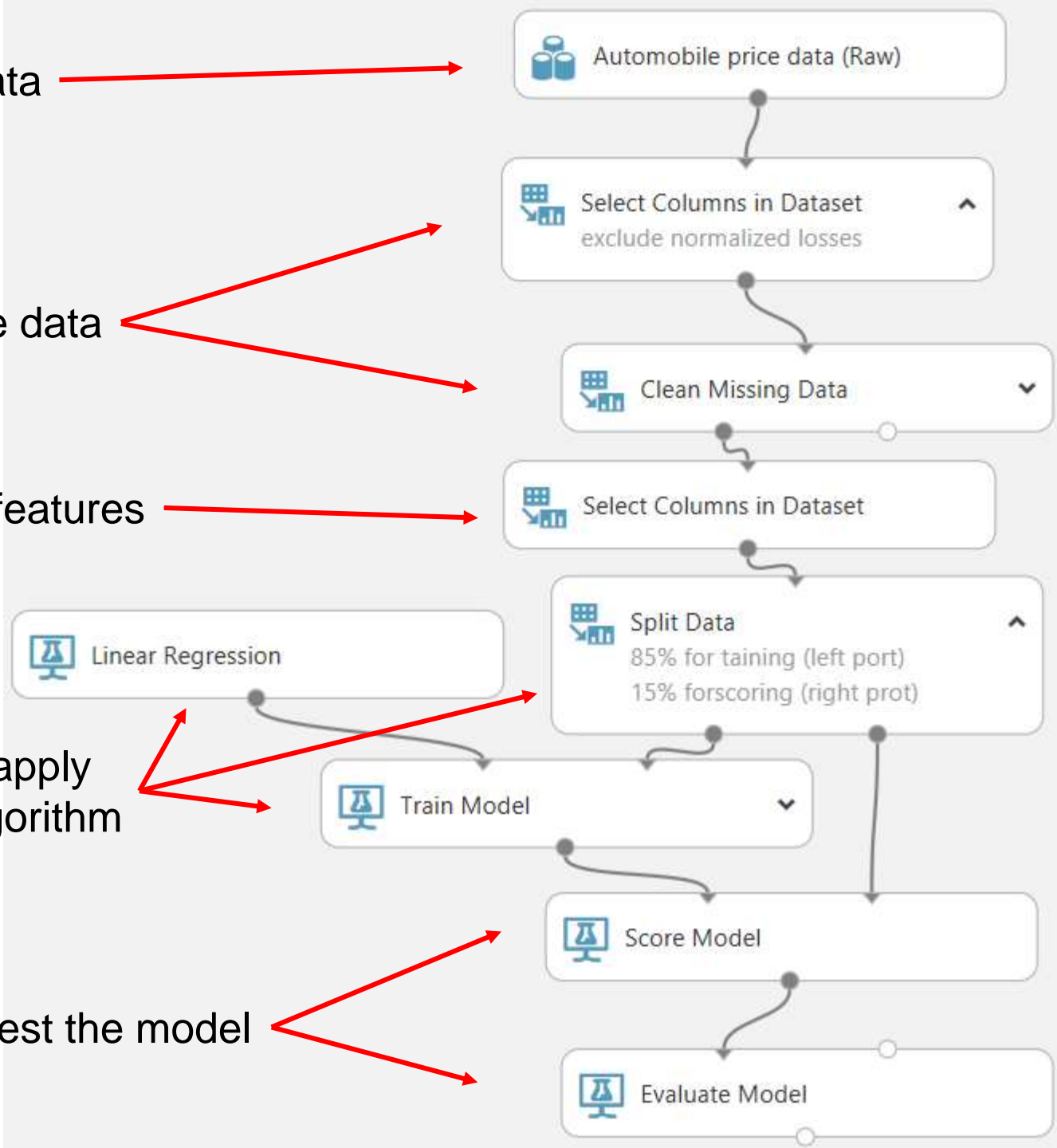
Choose and apply a learning algorithm

 Train Model

 Score Model

Score and test the model

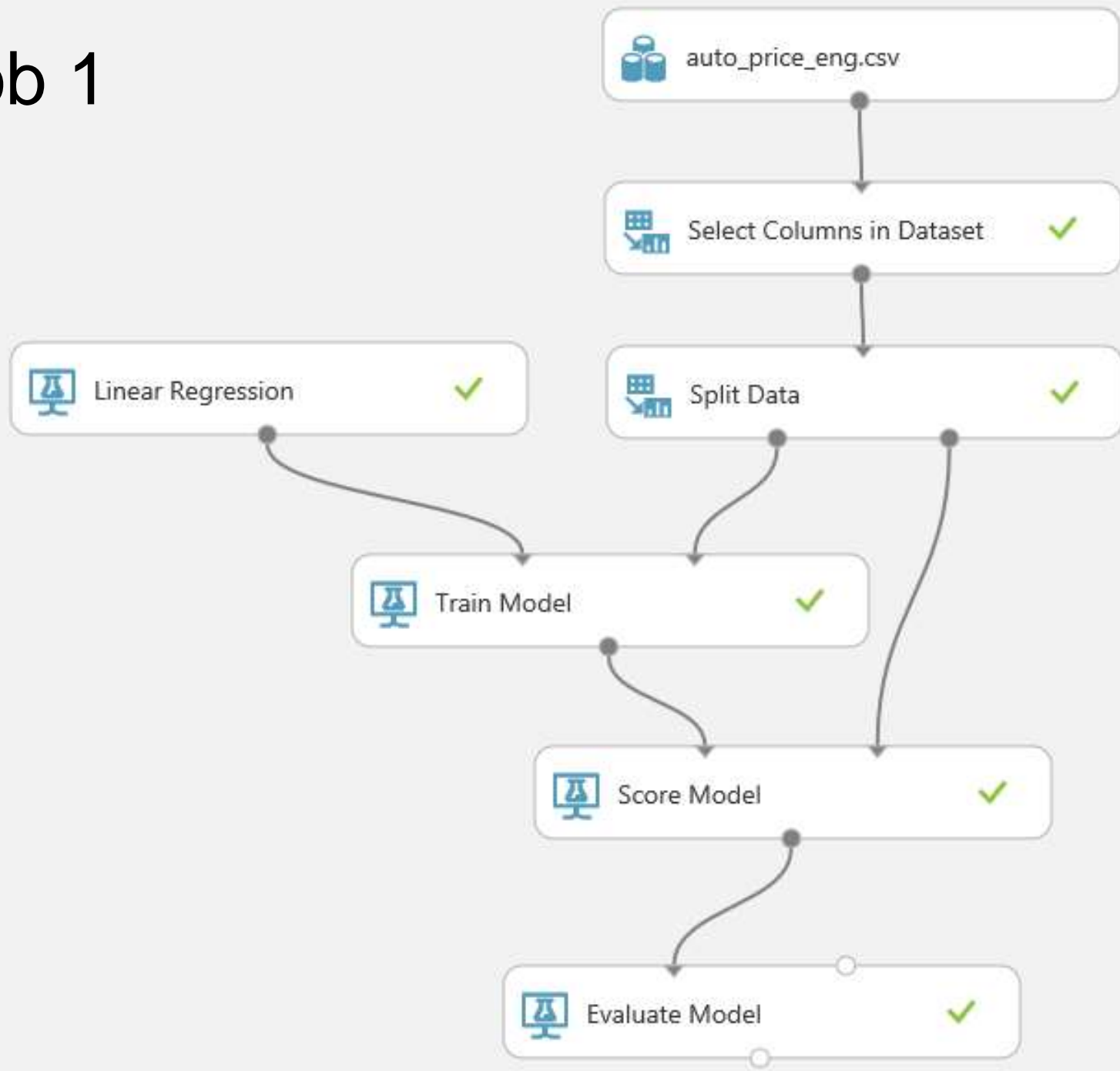
 Evaluate Model



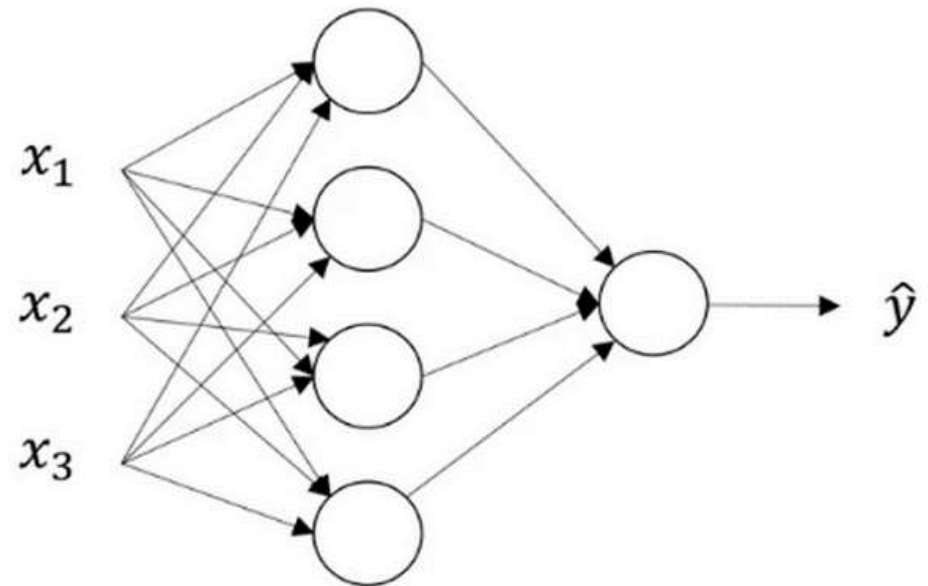
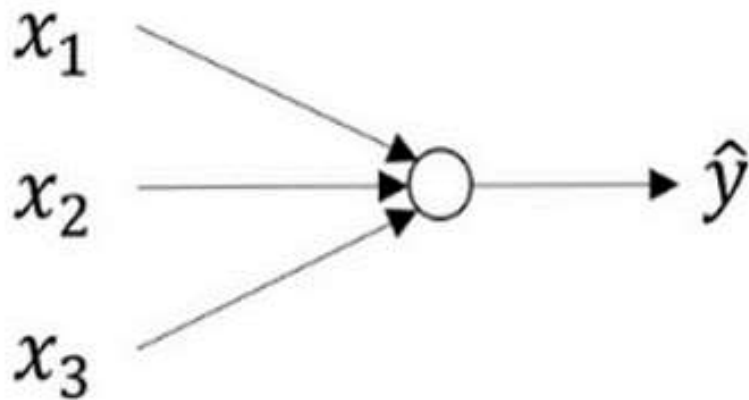
Prepare your data

- Example data (data for exercise.xlsx):
download from
<http://cms.sjtu.edu.cn/download.html>
- Convert it into two *.csv file (change the Chinese column name into Eng)
- Load the *.csv file to Azure Machine Learning Studio

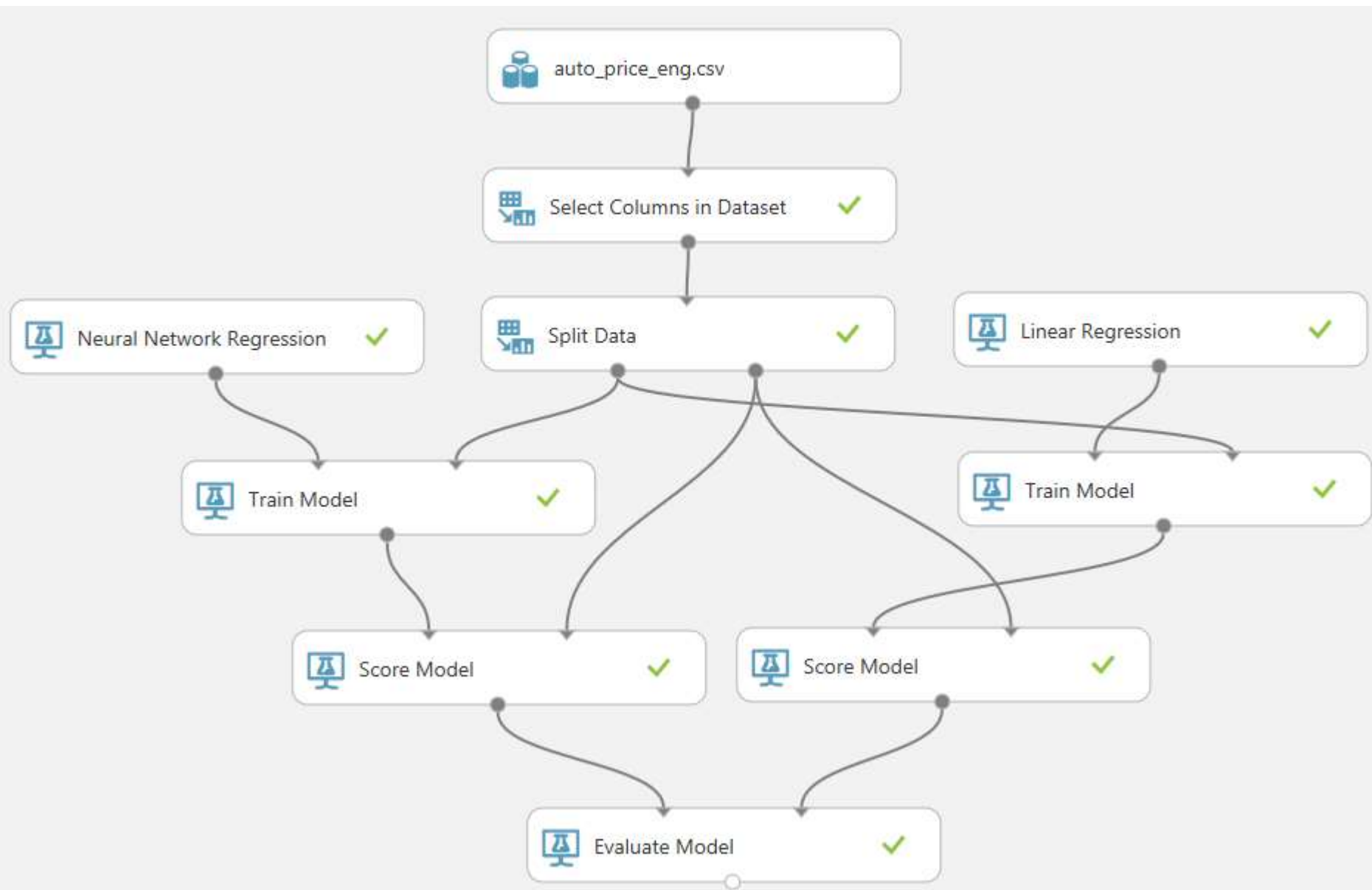
Job 1



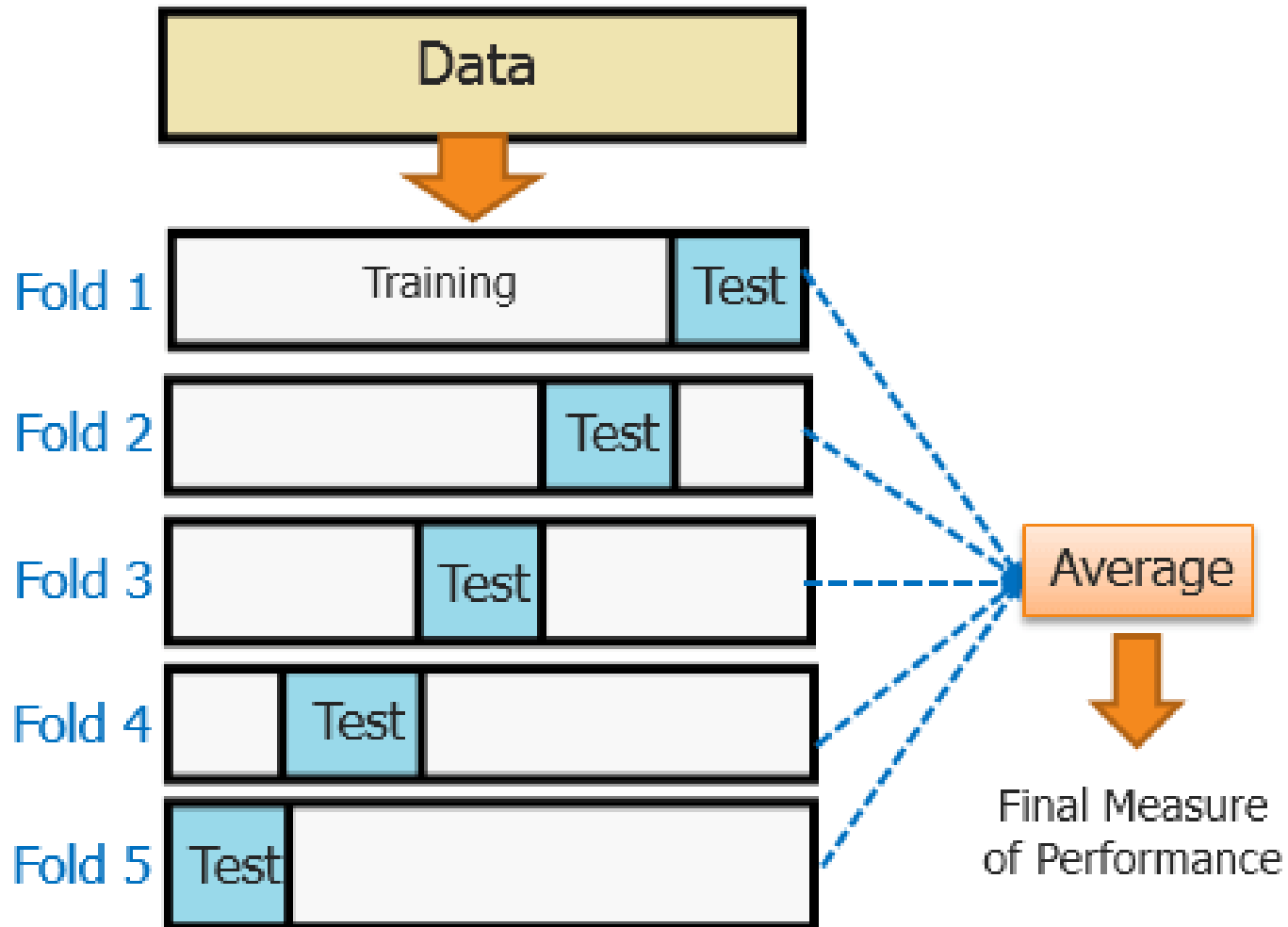
linear regression vs. neural network regression



nn regression vs. linear regression

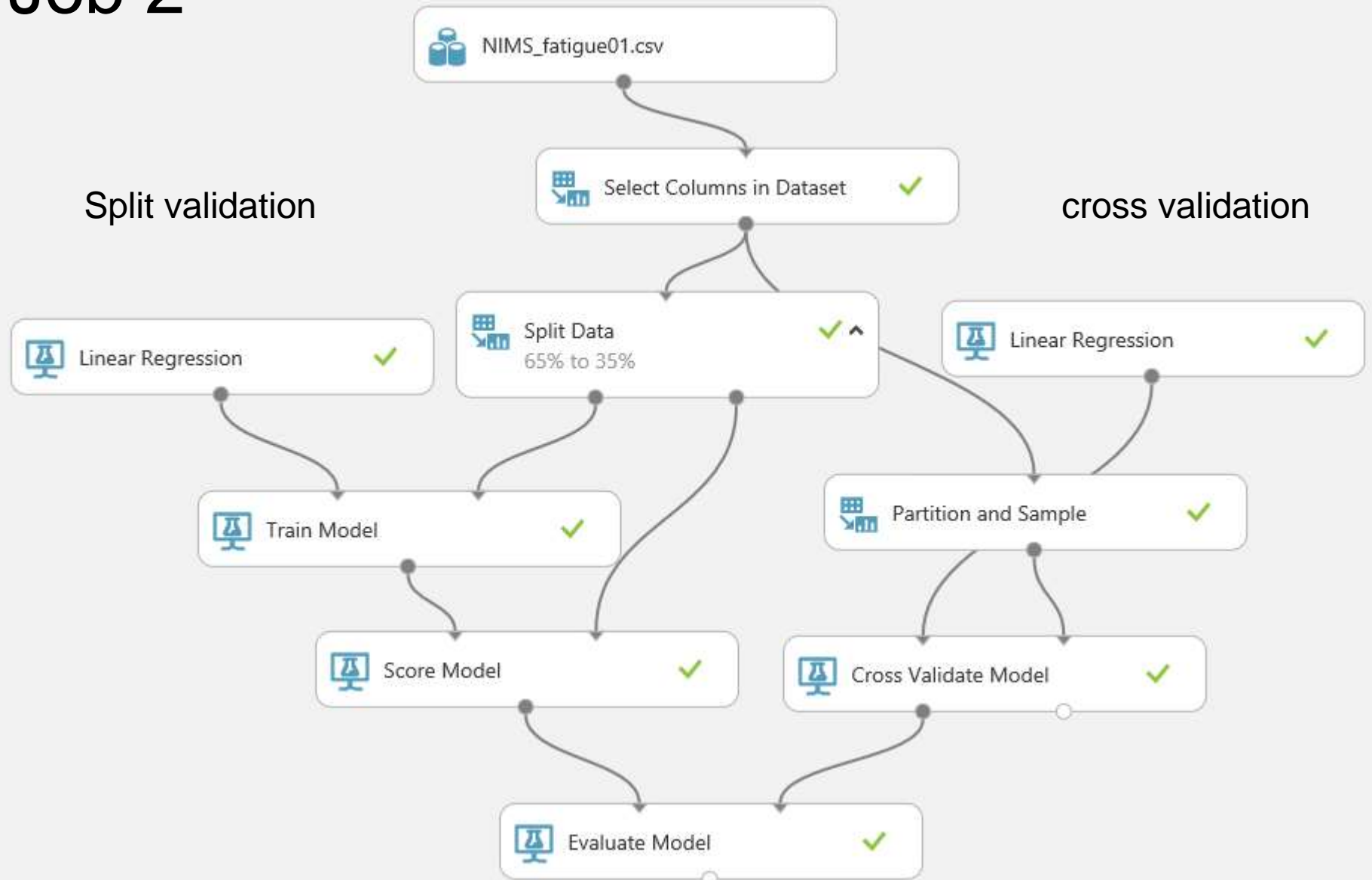


N fold cross validation(交叉验证)



fatigue strength predictor

Job 2



Classroom lab II: Magpie

- We need Java.
- Check your Java: **java -version** (command line, Win: cmd or PowerShell; Mac: terminal)

Windows PowerShell

```
Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

PS D:\Users\Lantingzh> java -version
java version "1.8.0_144"
Java(TM) SE Runtime Environment (build 1.8.0_144-b01)
Java HotSpot(TM) 64-Bit Server VM (build 25.144-b01, mixed mode)
PS D:\Users\Lantingzh> █
```

Install a correct version of Java

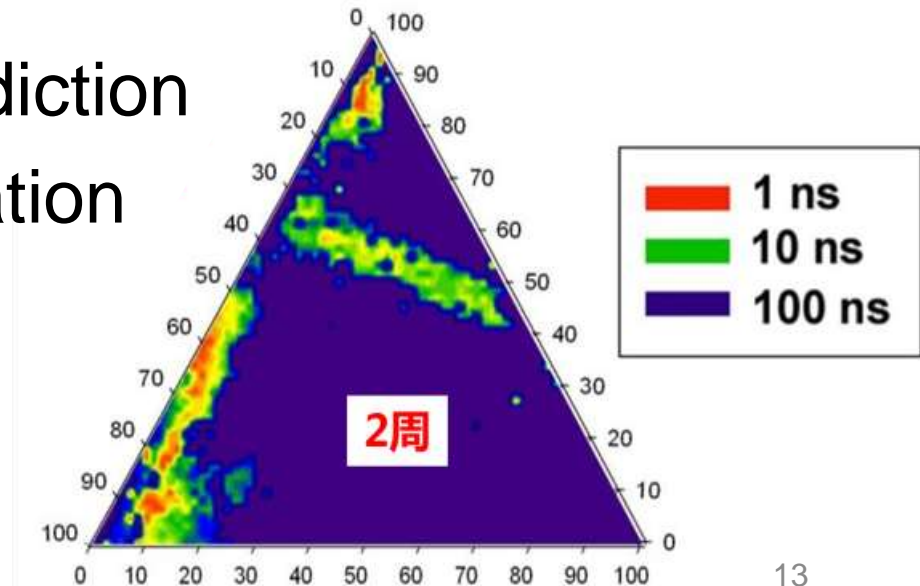
- For Win, **Java SE Runtime Environment 8u211** is needed. Visit <http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>
- For Mac, **Java SE Development Kit 8u211** is needed. Visit <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
don't download and install **Java SE 10.0.1**
- In case a wrong version is installed, to uninstall: <https://docs.oracle.com/javase/10/install/installation-jdk-and-jre-macos.htm#JSJIG-GUID-F9183C70-2E96-40F4-9104-F3814A5A331F>

课堂实验内容

- Use Magpie code (42 MB) (guide and code from <http://cms.sjtu.edu.cn/download.html>) to predict amorphous formation in a ternary system
- Walk through the code
- Predict Sb-Te-Ge system using supplementary data

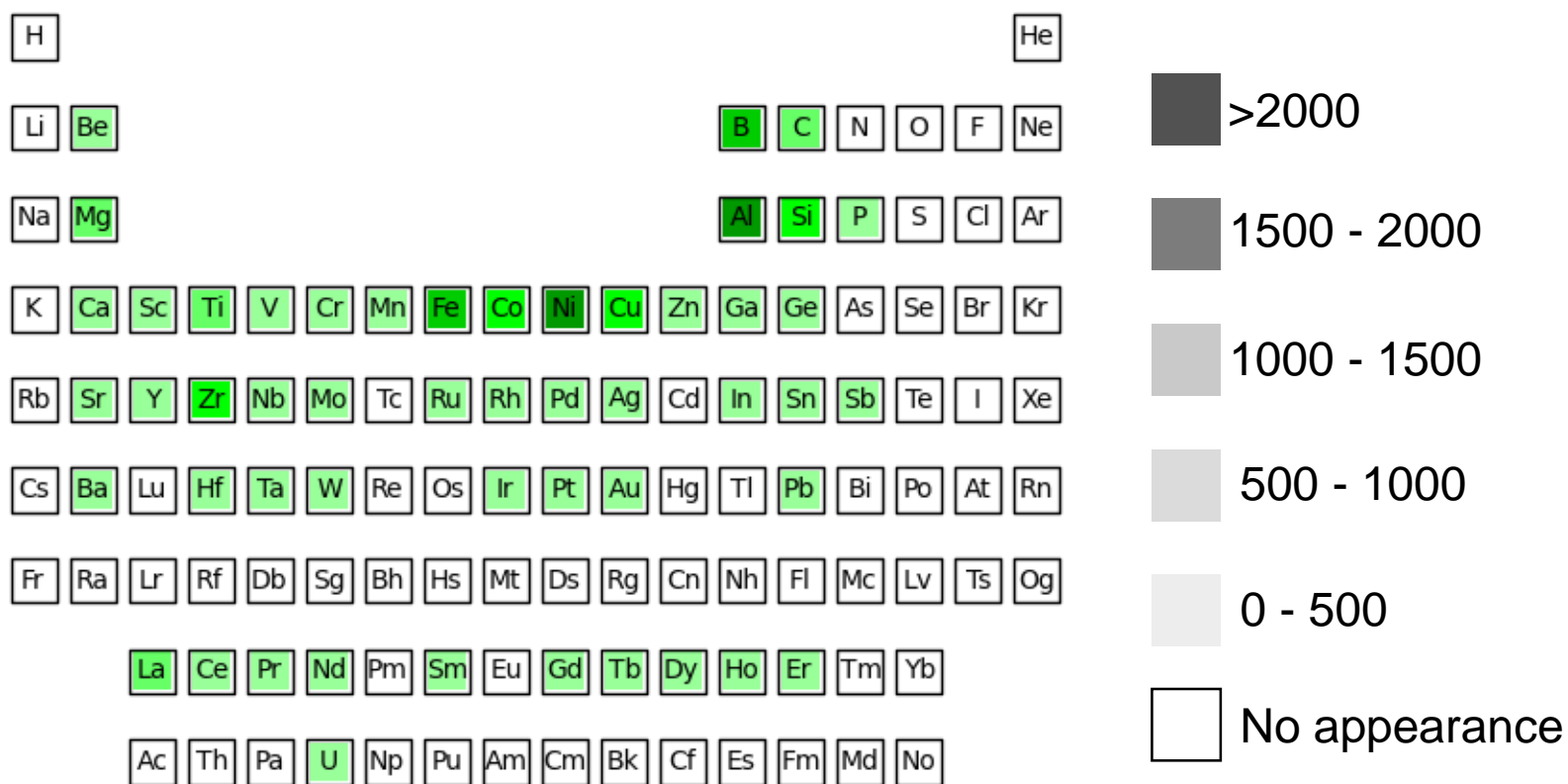
课堂实验内容

- Supplementary data of the Sb-Te-Ge is in the datasets/supplement folder in both *.csv and *.data forms, which is derived from the following experiment result
- Append the Sb-Te-Ge data to the glass.data dataset
- Write code to make prediction
- Use 10-fold cross-validation



原始数据中 各元素的出现频率： Al,Ni,Fe,B,Si,Mg,Zr,C,P ...

Abundance of elements in the data set



Course report (due Jun. 28)

- <ftp://public.sjtu.edu.cn>, user: lantingzh, code: public
- Upload file to \upload\cms2019\class1 or class2, upload pdf file only
- Name your file as 学号+姓名.pdf
- To predict a system using Magpie code (eg. Zr-Cu-Ni, Fe-Si-B, Sb-Te-Ge etc)
- Plot the predicted distribution in the composition space by certain means

Outline of course report (due Jun. 28)

- Title of your report (eg. Amorphous formation prediction in the x-x-x system)
- Your name and ID number
- Purpose of your computer experiment
- Code used in your work (the content in your *.in file)
- Method and software used to plot the predicted distribution
- Your output figure