Daedalus Python Workshop

L8: Machine Learning Basics

Alexander J. Peterson Santos & Peter Geldner

Garching, 18. December 2025

What is machine learning?

In short, algorithms that can be trained using a set of data in order to make generalizations/predictions about new, unseen data that falls into a similar distribution.

Supervised learning: Learn to make predictions on **labelled** data

- Predict a house price based on learned knowledge of other house-price pairs

Unsupervised learning: Find **patterns** in unlabelled data

- Group people into personality groups based on their answers to a survey

Topics

- Overview of machine learning
- Classification tasks
  - Logistic Regression model
  - k-Nearest Neighbor model
- Regression tasks
  - Linear Regression model

Supervised learning

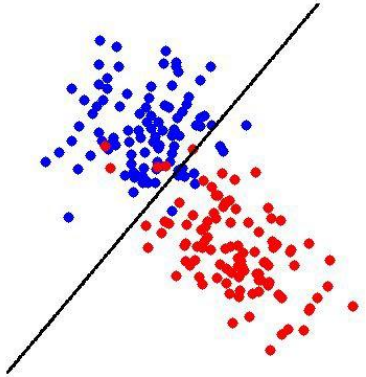
Machine Learning

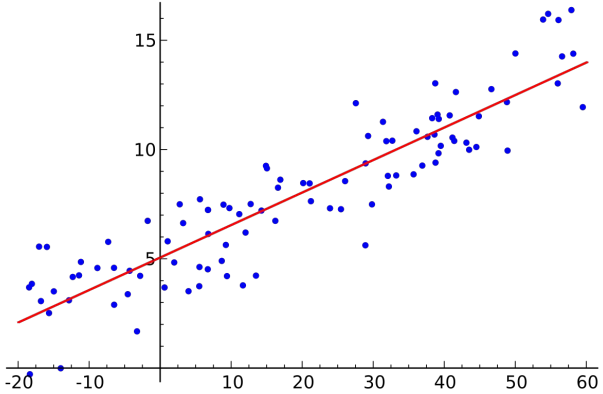
Supervised Learning

Unsupervised Learning

Classification
Discrete target vals

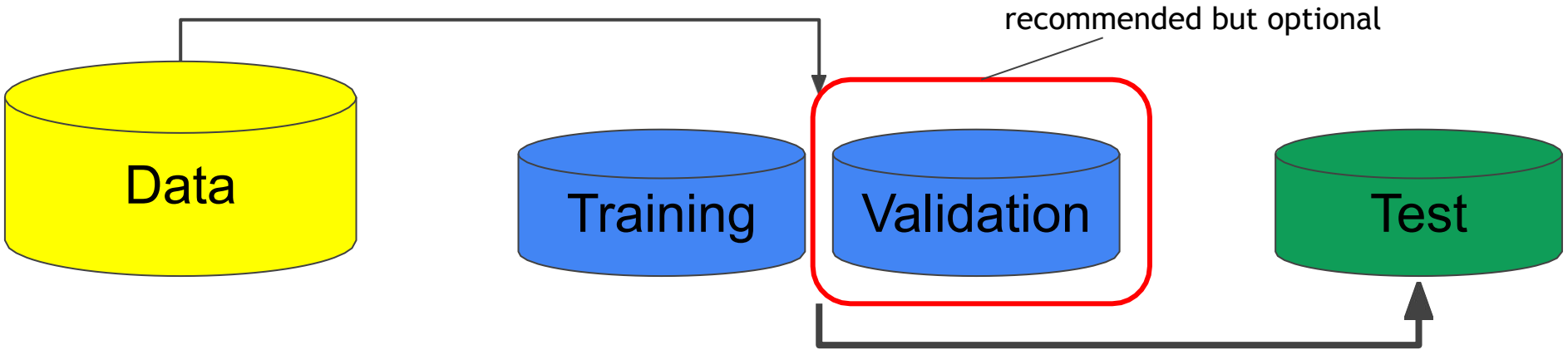
Regression
Continuous target vals





Supervised learning: train-test split

GOAL: Our ML is able to **generalize** on completely new data points!



Packages required for this lesson

You will need **numpy**, **pandas, matplotlib, scikit-learn** installed to your virtual environment.

Windows:

```
.\.venv\Scripts\Activate
pip install numpy pandas matplotlib scikit-learn
```

MacOS or Linux:

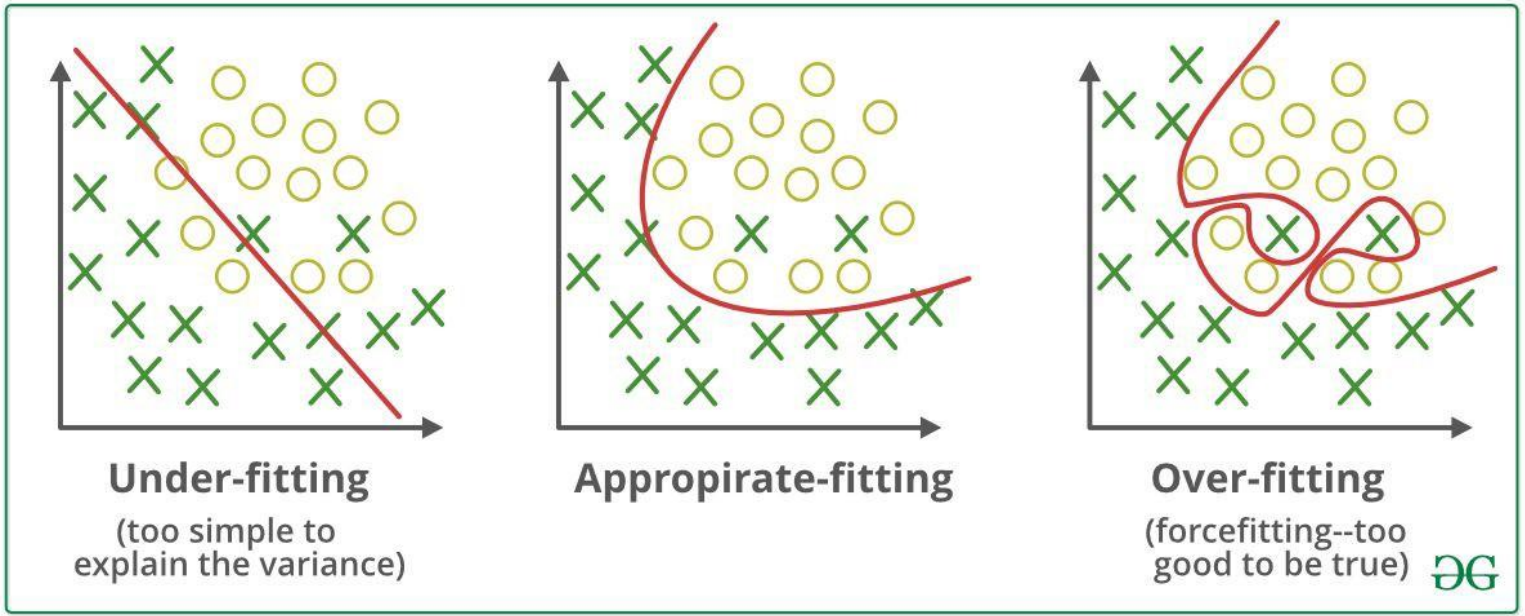
```
.venv/bin/activate
pip install numpy pandas matplotlib scikit-learn
```

Activate your virtual environment within the Jupyter Notebook in VSCode.

Let’s go to our Jupyter Notebook…

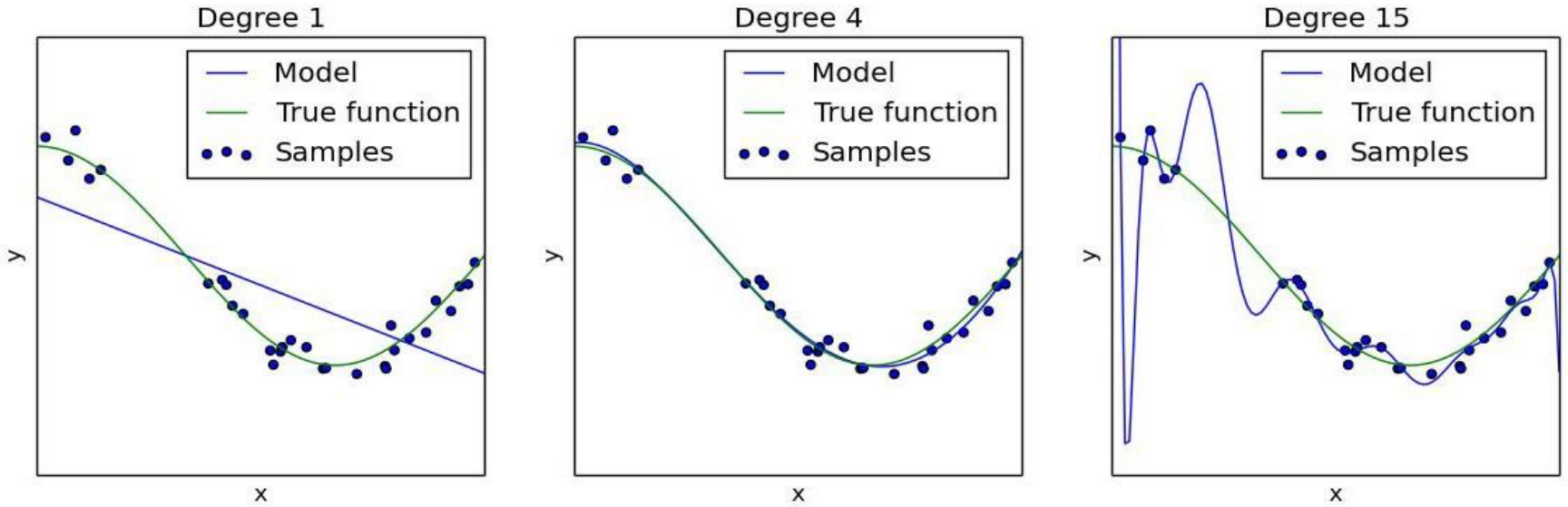
A word on overfitting / underfitting

Underfitting and overfitting



Overfitting in a binary classification model

Underfitting and overfitting



Overfitting in a regression model