This module is an overview of machine learning process and the traits in common of most learning algorithms.

1. **Getting to Know Machine Learning-** understand the basic ideas behind machine learning.

**Home Reading-** Aurélien Géron (2017), Hands-on Machine Learning with Scikit-Learn & TensorFlow, O’Reilly.

Chapter 1 one is available in pdf at: <https://drive.google.com/file/d/1OGAMM39DD-nEa8EpeM9pwfpGqVwCWf5n/view?usp=sharing> or you can read the whole book online from the campus library account.

* The Machine Learning Landscape page 3 - 14
* Main Challenges of Training Data page 22 - 31

Make lots of notes around the following concepts and how the relate to each other, diagrams help:

* Machine learning
* Training
* Data set (Training set, Validation set, Test set)
* General workflow of machine learning algorithms (Diagram)
* Overfitting
* Supervised Learning
* Underfitting
* Classification Task
* Regression Task

1. **Framework for machine learning:**

[The scikit framework technology stack](https://live.staticflickr.com/65535/48057247952_f9e106cb4a_b.jpg)

1. **Use SciKit API to train a linear regression:**

Linear regression with a library:

<https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html#sphx-glr-auto-examples-linear-model-plot-ols-py>

Here is the same example loaded into a notebook on colab [See example of regression](https://colab.research.google.com/drive/1iRjNJn5vsH9bFbU3v3sunhkmqd1Y64D5?usp=sharing)  make your own copy and add your own comments.

Here are some examples of different types of regressions from scikit API:

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>

Do your own notebook now load the [Boston data set](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html#sklearn.datasets.load_boston) and train your own regression, plot it afterwards

1. **Finding Data sets:**

Finding and using a dataset

UC Irvin <https://archive.ics.uci.edu/ml/index.php>

Data set from <https://index.okfn.org/place/>

UN <http://data.un.org/>

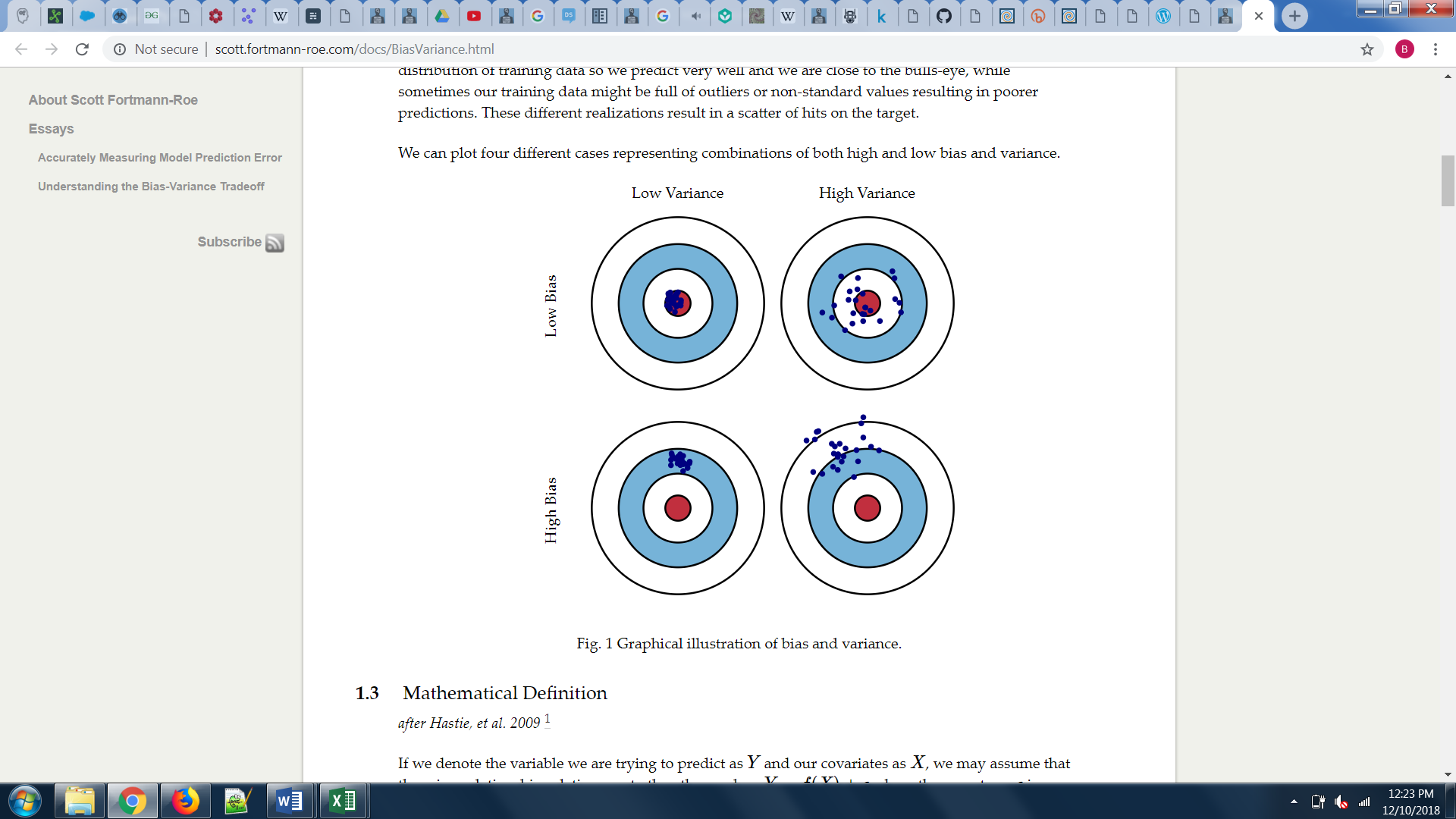
World bank <https://data.worldbank.org/>

Why is it necessary to split Train and Test

Find a regression dataset, load it and clean it using pandas or python and train a model with it.

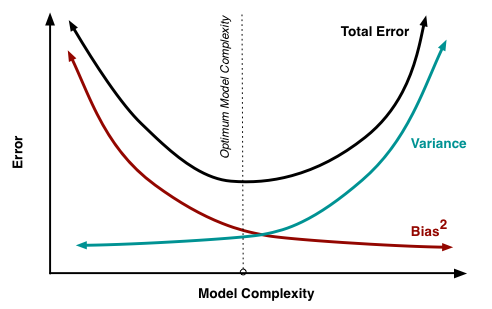
1. **Overfitting and Underfitting:**

See example in the file (what it looks like in data): <https://drive.google.com/file/d/1hz0cQsTNF2BW-iJrEqWIKvM1H8KhWc2a/view?usp=sharing>

Variance and bias <http://scott.fortmann-roe.com/docs/BiasVariance.html>

Learning to read the errors (Bias and Overfitting)

Training error



Cross-validation error “[The purpose of cross-validation is model checking, not model building](https://stats.stackexchange.com/questions/52274/how-to-choose-a-predictive-model-after-k-fold-cross-validation).”