

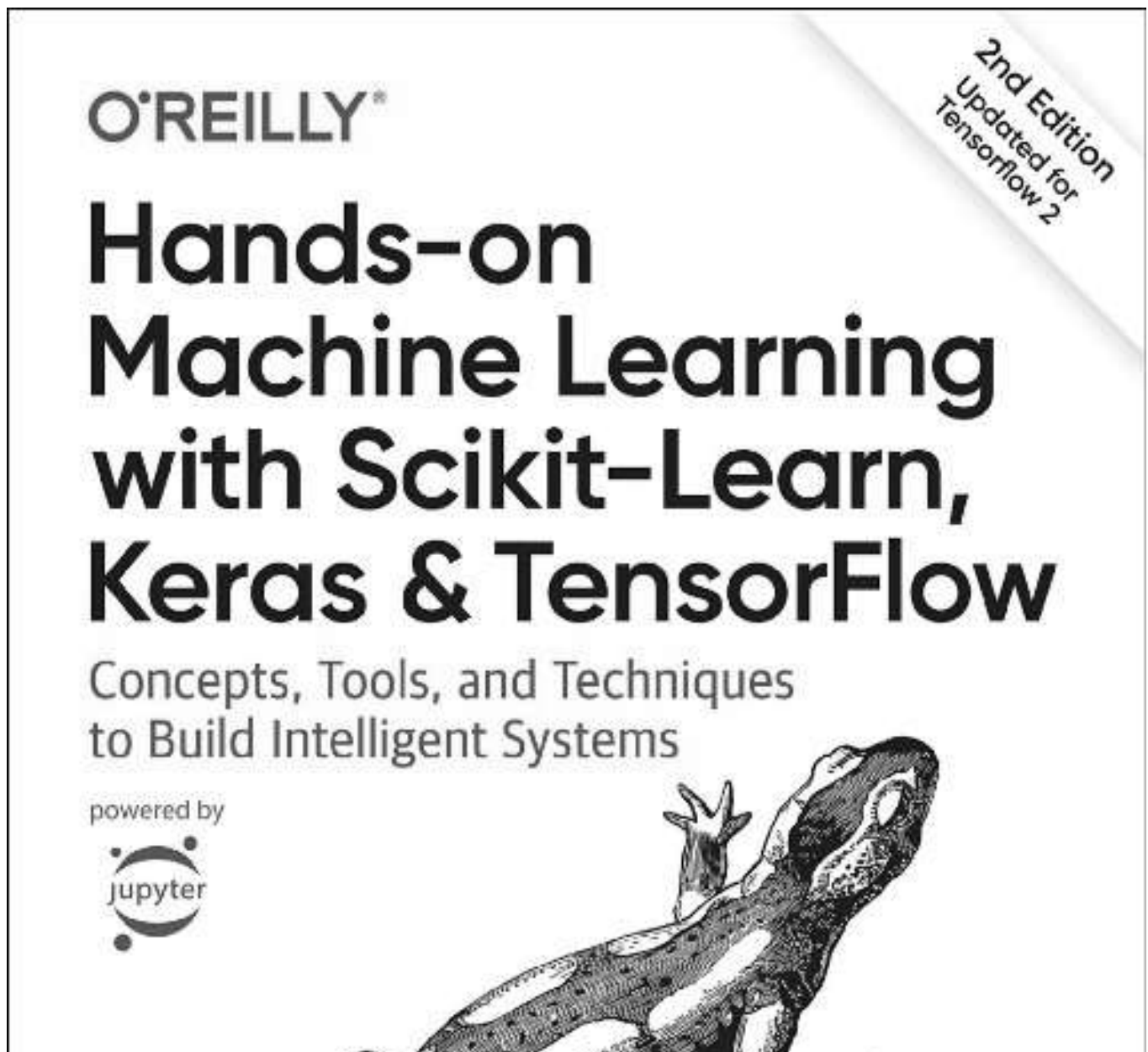
Hands-On ML, Edition 2 by Aerolien Geron, Summary Part 1



Aaron Ma [Follow](#)

Feb 7 · 4 min read

This is the first part of the 2 part series by Aaron Ma on the summary of Hands-On ML.





Pinterest image. Copyright Aerolien Geron and respective authors.

So you may have come across the image on the left: Hands-on ML with Scikit-Learn, Keras & TensorFlow by Aurelien Geron. I've just finished reading his book, and I've decided to create a neat little summary of his book.

. . .

I. The Fundamentals of Machine Learning

This book is divided into 2 sections. This is the 1st section's summary. You can find the 2nd section's summary [here](#). The 1st section covers the following:

- What ML is, the difference between Traditional Learning and ML, etc.
- Typical steps in training ML models
- Training ML models
- Optimizing a loss function
- Preparing data
- Engineering features
- One-hot encoding

- Most common ML models
- Multioutput classification & Multilabel classification
- Validating models

Chapter 1 — The Machine Learning Landscape

Welcome to the exciting world of machine learning! What is machine learning? Machine learning is the field of study of programming computers that learns from experience. In ML, there are three types: Supervised, Unsupervised, and Reinforcement learning. Supervised learning learns from data, and when predicting, it predicts the label(y). Unsupervised learning learns from data and clusters them together based on features(e.g., 100 books, 200 pens. The unsupervised learning algorithm will sort 100 publications in 1 place and 200 pens in 1 spot). Finally, Reinforcement learning learns from experience, using reward-based knowledge. Don't forget to click here for my previous talk on Reinforcement learning, where'll you'll build a self-driving car in a simulated environment! This chapter also covers the difference between online and offline learning. Online learning is what YouTube uses. Every once in a while, the data and models get retrained. Vice versa, offline learning is training the model once, so it won't be able to adapt to trends. A great way to learn these stuff is here.

. . .

Chapter 2 — End-to-End Machine Learning Project

This chapter builds a hands-on project using California Housing Price Dataset. This chapter emphasizes the Machine Learning workflow:

- 1. Get the data
- 2. Data analysis (remove useless features, plot graphs, etc.)
- 3. Select an algorithm
- 4. Create a model.

- 5. Evaluate the model's performance.
- 6. Fine-tune the models and combine them into a solution.
- 7. Present your solution.
- 8. Deploy and monitor your solution.

• • •

Chapter 3 — Classification

In this chapter, we start to create supervised ML models. It talks about different performance metrics:

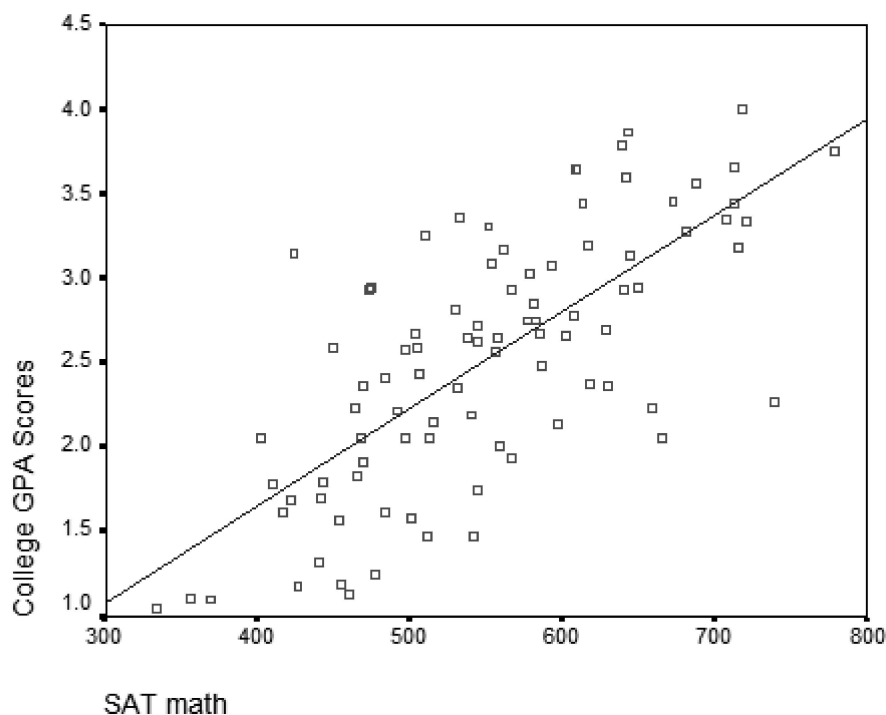
- Use Cross-Validation (K-Folds)
- Confusion Matrix
- Metric Options
- Precision
- Recall
- F1 score
- AUC/ROC

This chapter also teaches some core ML concepts like One-Hot encoding, where one value is [1], and the rest is [0], forming matrixes.

• • •

Chapter 4 — Training Models

This chapter teaches Linear Regression & Gradient Descent. Linear regression is a direct approach:



An image from Google Images.

As you can see, Linear Regression finds the best line of fit.

Gradient Descent, on the other hand, optimizes in shape like a bowl, taking small steps iteratively to the lowest part of the bowl.

. . .

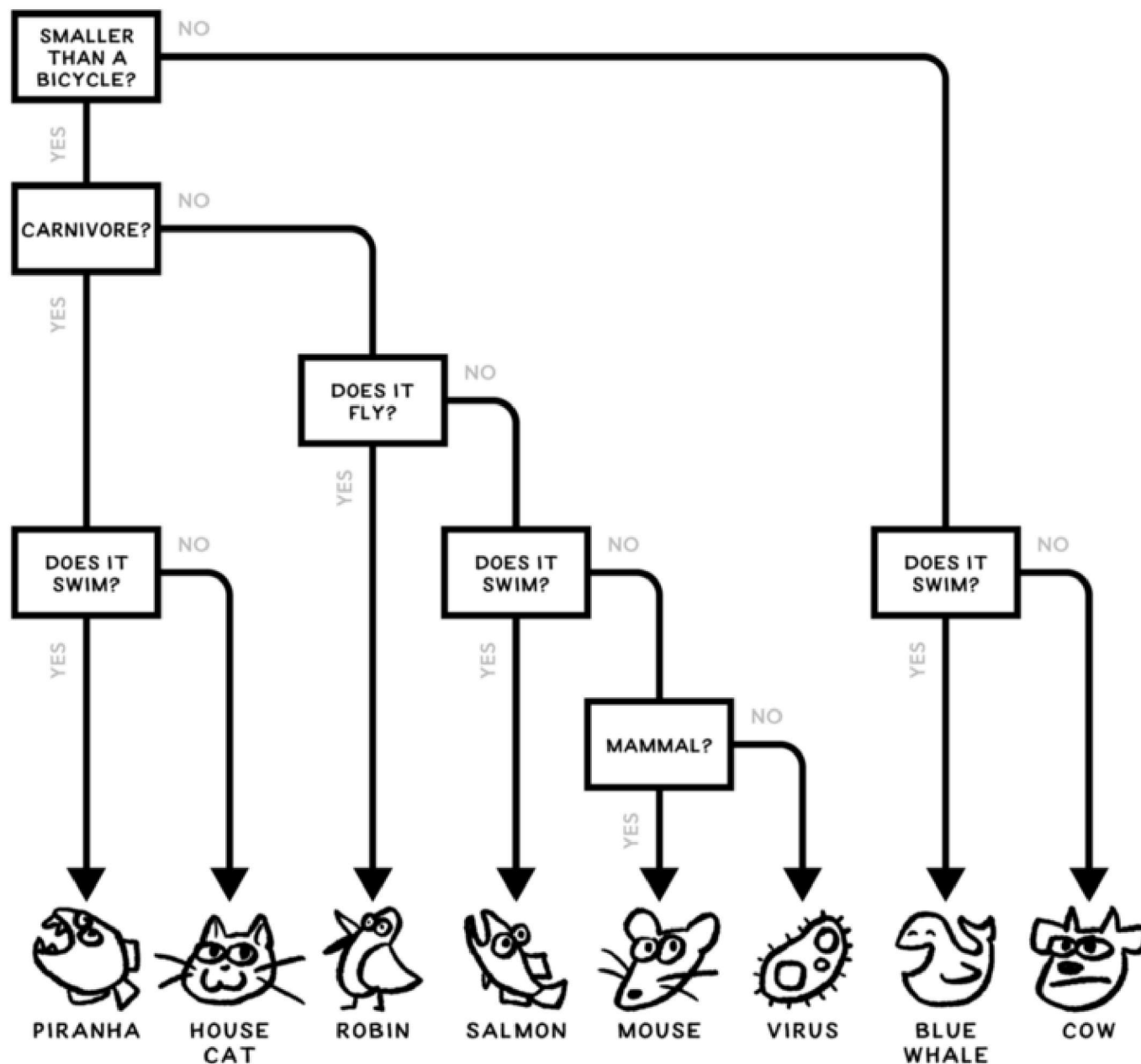
Chapter 5 — Support Vector Machines

Support vector machines(SVM) can be used for both regression & classification. It is formally defined as a “separating hyperplane” (drawing three lines).

. . .

Chapter 6 — Decision Trees

Decision trees are the most basic data structure of programming. Here's an example of one classifying different animals:



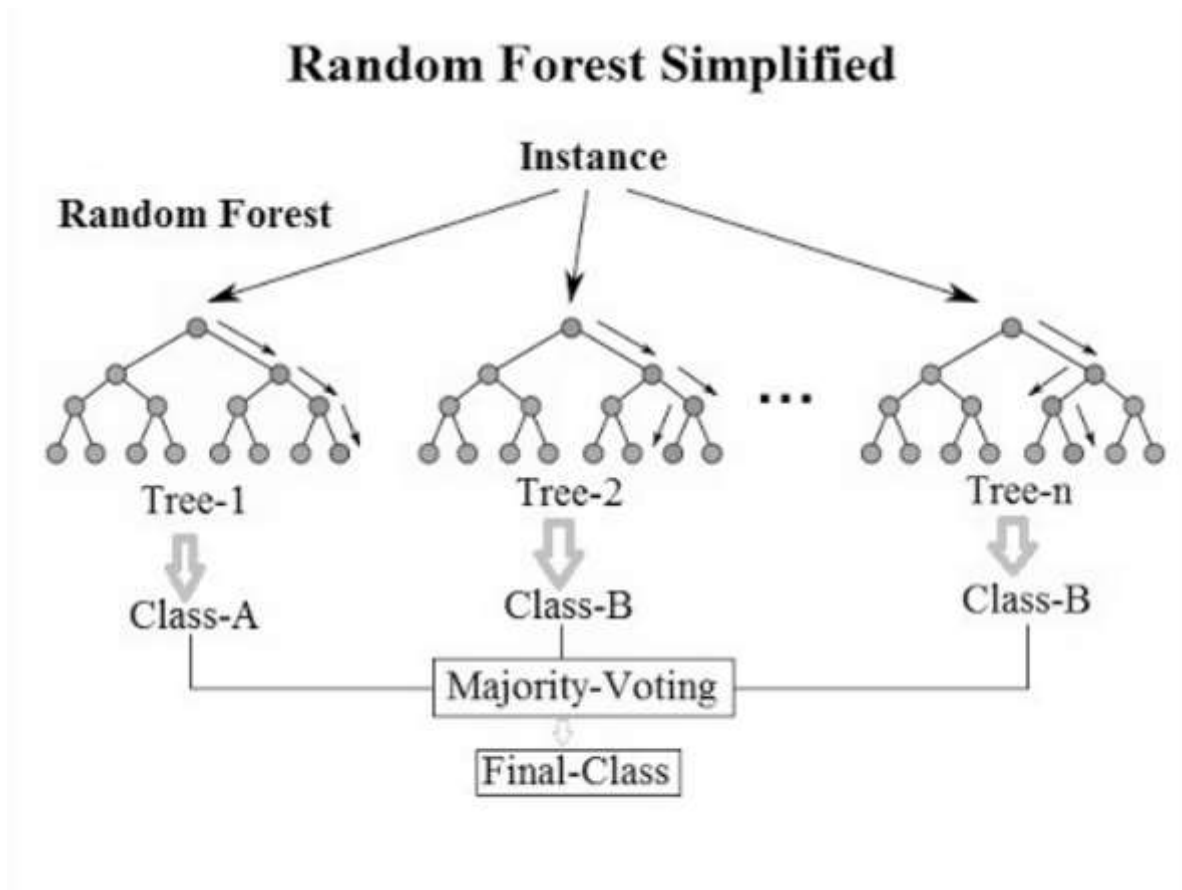
Copyright Brilliant's computer science course.

. . .

Chapter 7 — Ensemble Learning & Random Forests

Ensemble learning is where multiple learners are trained to solve the given ML problem and combine them.

Random forests are an ensemble learning method for classification and regression by creating decision trees:



Copyright miro.medium.com

. . .

Chapter 8 — Dimensionality Techniques

This chapter focuses on techniques to overcome the difficulty of high dimensional data. These datasets often cannot be held in memory together, as this demands a considerable

cost of computational power. Dimensionality techniques do not affect model performance. Here's an example:

If you worked at YouTube, you would have trillions of user's data for each training instance. Not only all of the features would make the training slow, but it will also confuse the model, which makes it harder to find an optimal solution in our model.

. . .

Chapter 9— Unsupervised Learning Techniques

This chapter focuses on clustering, K-means, and more! Read this on unsupervised learning techniques. Let's take a look at the techniques introduced in the book:

- Clustering

Clustering is where the ML algorithm will notice similarities in a dataset and cluster them together, where they belong. A sample of clustering in action is detecting cats vs. dogs.

- Anomaly detection

The objective here is to learn what “normal” instances are and use it to detect “unnormal” instances. A great example of this is skin cancer detection.

- Density estimation

This is the task of estimating the *probability density function(PDF)* of the random process that generated the dataset.

. . .

Thanks for reading! Don't forget to stay tuned for the next part of this article where I'll summarize part 2 of this book. Bye! — Aaron Ma

Machine Learning

Education

Aaron Ma

[About](#) [Help](#) [Legal](#)

Get the Medium app

