**The Engineers Guide to Machine Learning**

**机器学习工程师指南**

The definitive resource for all things Machine Learning in the galaxy

机器学习相关知识的权威资源

Machine learning is difficult and there is a lot going on. During a project earlier this month I was doing some simple data processing and I could not for the life of me recall the name of a bi-variate data exploration technique I read about a few months back. It was frustrating not to have a source of information right at my finger tips that I could easily get to.

机器学习是非常困难的，因为要掌握多方面的知识。在本月的早些时候进行的一个项目中，我在处理一些简单的数据时，我无论如何也回想不起一个双变量探索技术的名称，尽管我在几个月前刚刚读过。令我沮丧的是，在我眼前并没有一个可以简单获得的信息资源。

After spending some time searching on the internet I wasn’t able to find anything that matched what I had in mind. So I've decided to create the engineers guide to machine learning. An all inclusive mind map with most, if not all concepts and methods that would be useful to new and experienced machine learning engineers.

在网上花费了很多时间进行查找后，也没有找到符合我想法的资料。所以，我决定创建一份机器学习的工程师指南。一份全面的思维导图，在其中，即使不是所有的概念和方法，但大部分都是对新手和有经验的机器学习工程师非常有用的。

**The engineers guide to machine learning is broken up in to 5 different Sections**

指南被分为5部分：

* Machine Learning Data Processing
* 机器学习数据处理
* Machine Learning Concepts
* 机器学习概念
* Machine Learning Process
* 机器学习过程
* Machine Learning Mathematics
* 机器学习数学运算
* Machine Learning Models
* 机器学习模型

Each of these sections will go in depth on topics that are included in it. I’m working on the bible now. Sections will be released every week and will be accompanied with a corresponding “Cheat Sheet” for easy reference.

这其中的每一部分将深入讨论它所包含的主题。我现在正在写这本“圣经”。每周都会发布一些章节，并附上相应的“速查表”，以便快速查找。

Here is a taste of the first section. The only thing you might need after I publish all the sections is a towel. Let me know if I should add anything!

这儿先尝一下第一部分。等我发布完所有的章节，你唯一需要的事情就是一条毛巾。让我知道，我应该需要添加什么东西！

**Machine Learning Data processing**

**机器学习数据处理**

**Data Types**

**数据类型**

* Nominal
* Ordinal
* Interval
* Ratio

**Data Exploration**

**数据挖掘**

* Variable Identification
* Uni-variate Analysis
* Bi-variate Analysis
* Multi-variate Analysis

**Feature Cleaning**

**特征清洗**

* Missing Values
* Special Values
* Outliers
* Obvious inconsistencies

**Feature Imputation**

* Hot-Deck（热卡法）
* Cold-Deck（冷卡法）
* Mean-substitution
* Regression（回归法）

**Feature Engineering**

**特征工程**

* Decomposition
* Dicretization
* Reframe Numerical Quantities
* Crossing

**Feature Selection**

**特征选取**

* Correlations
* Dimensionality Reduction
* Importance

**Feature Encoding**

**特征编码**

* Label Encoding
* 标签编码
* One hot Encoding
* 独热码

**Feature Normalization**

**特征归一化**

* Re-scaling
* Standardization
* Scaling to unit Length

**Dataset Construction**

**数据集构建**

* Training Dataset
* Test Dataset
* Validation Dataset
* Cross validation