ds_book

Igor Veksler

12/4/22

Table of contents

Preface

This book will cover data science concepts. This will contain longer form content from the ds_skills repository

1 Introduction & Motivations

I have always been interested in how can somebody make a meaningful contribution to Aritifical Intelligence without a formal education or a phd. ("playa hatin degree") This spans the spectrum of being a practitioner in industry to making theoretical advancements via a paper. I originally wrote about this topic in the 2018 post Bootstrapping a Modern Data Science Education while at DataRobot. My views have updated and this provides deeper content along this path. This repository will contain longer form content from the ds_skills repository

The book Introduction to Probability for Data Science by Stnaley H. Chan provides a glimpse into the future of publishing technical content. This contains longer form text explanations as well as code samples and video lectures. This provides all the modalities of learning on behalf of the student.

We are on the precepice of a golden age of education as now so much of the worlds content from the leading minds is readily accessible to anyone for free.

We live in an AWESOME age of enlightenment. Oh, you wanna learn about SVD? Have (**luis_likes_math?**) break it down for you[1]. Or have (**3blue1brown?**) show you how to bend space with your mind (& linear algebra) [2]. Or just attend the whole MIT course [3].We're incredibly blessed.

— Jay Alammar ((**JayAlammar?**)) February 5, 2021

2 Summary

This book will cover data science concepts. This will contain longer form content from the ds_skills repository

Part I

test

3

Mathematical Foundations

Algebra is arithmetic that includes non-numerical entites like x. I.E. $2 * x + 5 = 25$

Linear algebra was developed to solve linear equations. "Solving for unknowns within a system of linear equations". If there is a non-linear transformation it doesn't fall within this definition. It can provide study for objects such as lines, planes and hyperplanes. It can be traced back to ancient civilizations.

Tensors are a machine learning generalization of vectors and matricies to any number of dimensions. I.E. An extra dimension for the number of observations in the MNIST dataset.

Code examples and notebooks for the content above: Intro to linear algebra - Jon Krohn Linear Algebra 2: Matrix Operations - Jon Krohn Data Science from scratch(linear algebra) - Joel Grus * Colab Notebook #todo: Add colab to repo

Classes: Mathematical Foundations of Machine Learning - Jon Krohn Essence of linear algebra - 3brown1blue

4 Optimization

Optimization is the study of finding the best output from the best combination of inputs.

Part II

Software_Engineering

Part III

According to wikipedia,

> *Domain knowledge is knowledge of a specific, specialized discipline or field, in contrast to general (or domain-independent) knowledge. The term is often used in reference to a more general discipline—for example, in describing a software engineer who has general knowledge of computer programming as well as domain knowledge about developing programs for a particular industry. People with domain knowledge are often regarded as specialists or experts in their field.* [1]

The definition I like is:

> *"In Datascience the term domain knowledge is used to refer to the general background knowledge of the field or environment to which the methods of data science are being applied."*

References:

- [1] Domain Knowledge - Wikipedia
- [2] https://corporatefinanceinstitute.com/resources/data-science/domain-knowledge-data-science/

5

Related to domain knowledge is knowledge of applications and use cases of ML. Elena Samuylova and the Evidently AI team meaintain a nice database of case studies of industry use cases:

- https://www.evidentlyai.com/ml-system-design

6

Placeholder for forecasting principles for the product book

Forecasting Principles

> Good forecasting is pretty boring

- Base rate
- Outside view
- etc.

References