# CSE 431

# Task 8 Report

# Summer 2023

*Abid Hossain*

*ID: 20301115*

**Submitted to: Annajiat Alim Rasel sir**

BRAC
UNIVERSITY

Inspiring Excellence

# Review of Stanford CS230: Deep Learning | Autumn 2018 | Lecture 8 - Career Advice / Reading Research Papers

I recently completed the Stanford CS230: Deep Learning course on YouTube, and I found it to be an exceptionally enriching experience. Personally, I consider it invaluable as it equipped me with essential skills that hold tremendous relevance in today's rapidly evolving world.

Foremost among the skills I acquired is a heightened proficiency in research techniques. This course offered insights into advanced methods of data preprocessing and preparation, which are fundamental in ensuring the robustness of any deep learning endeavor. Additionally, I gained a comprehensive understanding of the diverse spectrum of neural network architectures and the art of discerning the most suitable one for a given task.

Equally significant is the competence I developed in the fine-tuning and optimization of models during the training process. This encompassed the nuanced selection of hyperparameters, including critical factors like learning rate and batch size. I also grasped the application of techniques such as early stopping and weight decay, which play pivotal roles in averting overfitting. This facet of the course enriched my comprehension of the underlying mathematics behind the algorithms in use, enabling more informed decisions during model optimization.

Beyond the technical aspects, the course emphasized the paramount importance of cultivating strong research skills for effective participation in deep learning research. This entails the ability

to proficiently locate and evaluate pertinent literature, as well as the skill to design and execute experiments meticulously and reproducibly. I gained a profound appreciation for the significance of clear communication of research findings, whether through scholarly writing or dynamic presentations at conferences and peer-reviewed publications.

In summation, my engagement with the Stanford CS230: Deep Learning course has not only broadened my technical expertise but has also instilled in me a profound appreciation for the holistic research process. This experience has fortified my readiness to navigate the ever-evolving landscape of deep learning with competence and insight.

## Review of Fall 1999 Sl 18.01 Lectures of Gilbert Strang on Linear Algebra

On December 13th, I accomplished the task of watching the entire course titled "Lectures of Gilbert Strand on Linear Algebra," which was accessible to all viewers via the YouTube platform. I must express that this experience was profoundly enriching and eye-opening. In conjunction with content from "3blue1brown," this course provided me with a novel and innovative perspective on the subject of Linear Algebra.

One particularly memorable aspect of the course was the insightful analogies drawn between the concepts of eigenvectors and eigenvalues and their applications in speech processing. This

unique connection illuminated the practical relevance of Linear Algebra in the realm of signal processing and added depth to my understanding.

Furthermore, I gained a heightened appreciation for the pervasive role of Linear Algebra in various fields, especially in the domain of engineering. The course underscored how many key concepts in optimization and integration are fundamentally rooted in Linear Algebra. This mathematical discipline plays a foundational role in the design and analysis of critical structures such as bridges and buildings, highlighting its real-world significance.

From a personal standpoint, the most enjoyable segment of the course revolved around the challenging puzzles that required determining whether a specific set of vectors existed within an arbitrary vector space. Additionally, exploring the intricate behaviors and properties of vector spaces, such as basis and dimensions, provided a stimulating intellectual exercise. This process fostered critical thinking and problem-solving skills, making the course both engaging and intellectually rewarding.

In summation, my completion of the "Lectures of Gilbert Strand on Linear Algebra" course marked a significant milestone in my pursuit of mathematical knowledge. The course not only broadened my horizons with innovative perspectives but also underscored the ubiquitous nature of Linear Algebra in diverse disciplines. The engaging puzzles and problem-solving exercises further solidified my grasp of this mathematical field and left me with a profound sense of accomplishment.

# Review of tasks completed in R

In my experience with R programming, I have delved into a wide range of topics, gaining proficiency in various aspects of this versatile language. Here is a more detailed breakdown of the R-related tasks I have undertaken:

1. Basic Syntax: I have developed a solid understanding of the basic syntax of R, including variable assignment, data types, control structures (such as loops and conditionals), and functions. This foundation has been essential for my work in data manipulation and analysis.

2. Data Manipulation: Data manipulation is a core skill in R, and I have honed my abilities in this area. I am well-versed in using functions from packages like `dplyr` and `tidyr` to efficiently filter, arrange, mutate, and summarize data frames. This allows me to transform raw data into meaningful insights.

3. Data Visualization: Visualization is a crucial part of data analysis, and I have explored various visualization techniques using R. I have utilized the powerful `ggplot2` package for creating customized and publication-quality plots, enabling me to communicate data-driven stories effectively.

4. Data Aggregation and Summarization: Aggregating and summarizing data is essential for generating meaningful statistics and insights. I have used R to perform operations like group-wise summarization, pivot tables, and descriptive statistics to extract valuable information from datasets.

5. Tidyverse: The Tidyverse is a game-changer in the R ecosystem, and I have harnessed its full potential. It consists of a collection of interconnected packages, including `ggplot2`, `dplyr`, and `tidyr`, which streamline data analysis workflows. Leveraging these packages has made my data manipulation and visualization tasks more efficient and coherent.

6. Data Import and Export: Working with data often begins with importing it into R. I have acquired skills in reading various file formats, such as CSV, Excel, and databases, into data frames. Additionally, I can export data frames to different formats for sharing and further analysis.

7. Vector and String Operations: R is exceptional for vectorized operations, and I have leveraged this feature to efficiently manipulate and analyze data. I am proficient in working with vectors and strings, performing tasks like concatenation, substring extraction, and vectorized mathematical operations.

8. GitHub: To showcase my R skills and projects, I have maintained a GitHub repository where I regularly upload my work. This portfolio includes detailed descriptions of the tasks I've

completed, code samples, project documentation, and visualizations. This allows potential collaborators or employers to review my work and assess my capabilities.

By combining these skills and experiences, I have become a proficient R programmer capable of handling diverse data-related tasks and providing valuable insights. You can explore my GitHub here to see examples of my work and gain a more comprehensive understanding of my capabilities in R programming.

https://github.com/abidbhay/CSE431/blob/main/submission8/R_learning.txt

## Review of OpenRefine

I have had the opportunity to work extensively with Google OpenRefine, a powerful data cleansing and wrangling tool that plays a pivotal role in data preparation for analysis and visualization. Let's delve deeper into the capabilities and features of OpenRefine:

1. Data Cleansing and Wrangling: Google OpenRefine is designed to assist users in the exploration, cleaning, and transformation of data from diverse sources. Its user-friendly interface makes it accessible to both beginners and experienced data professionals. It excels in tasks such as standardizing data formats, handling outliers, and resolving inconsistencies within datasets.

2. Data Source Compatibility: OpenRefine is exceptionally versatile in terms of data source compatibility. It can seamlessly process data from a wide range of formats, including CSV, Excel (both XLSX and XLS), JSON, TSV, and more. This flexibility ensures that it can accommodate various data structures commonly encountered in data analysis.

3. Large Dataset Handling: One of the standout features of OpenRefine is its ability to efficiently manage and manipulate large datasets. It can comfortably handle datasets comprising millions of rows and columns, making it an indispensable tool for data scientists and analysts dealing with extensive data volumes. Its performance remains robust even when dealing with substantial data sizes.

4. Data Transformation: OpenRefine offers a comprehensive suite of data transformation capabilities. Users can perform operations such as filtering, sorting, aggregating, and reshaping data with ease. This allows for the seamless preparation of data for downstream analysis, making it a preferred choice for data preprocessing tasks.

5. Data Quality Enhancement: OpenRefine goes beyond basic data manipulation by incorporating advanced algorithms for error detection and correction. It can identify and rectify common data issues, including spelling mistakes, duplicates, missing values, and structural inconsistencies. This data cleaning capability is crucial for ensuring the accuracy and reliability of analysis results.

6. Extensible and Customizable: OpenRefine is highly extensible and customizable through the use of plugins and scripting. Users can develop custom transformations and operations tailored to their specific data processing requirements. This extensibility adds a layer of adaptability to the tool, allowing it to cater to a wide array of data challenges.

7. Community Support: OpenRefine benefits from an active and supportive user community. This community actively contributes plugins, extensions, and documentation, making it easier for users to extend the tool's functionality and troubleshoot any issues they may encounter.

In conclusion, Google OpenRefine, formerly known as Google Refine, is an invaluable asset in the data professional's toolkit. Its versatility, scalability, and data-cleaning prowess make it a go-to solution for data preparation tasks in various industries and domains. Whether you're dealing with small datasets or massive data volumes, OpenRefine empowers you to wrangle, clean, and transform your data efficiently and effectively, ultimately enhancing the quality and reliability of your analytical insights.

I am providing and the tasks and additional work I have done in the github link below:

https://github.com/abidbhay/CSE431/tree/main/submission8