

On Machine Learning Interviews

Zeeshan Hooda¹²

27th November, 2023

¹Department of EECS, UC Berkeley

²Sparse Labs Inc.

Abstract

A collection of notes, learnings, exercises, and solutions developed from working through Chip Huyen's *Machine Learning Interviews*[\[1\]](#). Sections map directly to their sections in the book unless otherwise specified. We will do our best to provide answers for as many exercise questions as possible.

For feedback, errata, and suggestions, reach us at zee@sparselabs.org. Copyright ©2023 Sparse Labs Inc., All Rights Reserved.

Contents

I	Overview	4
1	Machine Learning Jobs	6
1.1	Different Machine Learning Roles	6
1.1.1	Working in Research vs Production	6
1.1.2	Research	7
1.1.2.1	Research vs Applied Research	7
1.1.2.2	Research Scientist vs Research Engineer	8
1.1.3	Production	8
1.1.3.1	Production Cycle	8
1.1.3.2	ML Engineer vs Software Engineer	10
1.1.3.3	ML Engineer vs Data Scientist	11
1.1.3.4	Other Technical Roles in ML Production	11
1.1.3.5	Understanding Roles and Titles	12
1.2	Types of Companies	12
1.2.1	Applications Companies vs Tooling Companies	12
1.2.2	Enterprise vs Consumer Products	12
1.2.3	Startups or Big Companies	12
2	Machine Learning Interview Process	13
2.1	Understanding the Interviewers' Mindset	14
2.1.1	What Companies Want From Candidates	14
2.1.1.1	Technical Skills	14
2.1.1.2	Non-technical Skills	14
2.1.1.3	What Exactly is Culture Fit?	14
2.1.1.4	Junior vs Senior Roles	14
2.1.1.5	Do I need a PhD to Work in ML?	14
2.1.2	How Companies Source Candidates	14
2.1.3	What Signals Companies Look for in Candidates	14
2.2	Interview Pipeline	14
2.2.1	Common Interview Formats	14
2.2.2	Alternative Interview Formats	14
2.2.3	Interviews at Big Companies vs at Small Companies	14
2.2.4	Interviews for Internships vs for Full-time Positions	14
2.3	Types of Questions	14

2.3.1	Behavioral Questions	14
2.3.1.1	Background and Resume	14
2.3.1.2	Interests	14
2.3.1.3	Communication	14
2.3.1.4	Personality	14
2.3.2	Questions to Ask Your Interviewers	14
2.3.3	Bad Interview Questions	14
2.4	Red Flags	14
2.5	Timeline	14
2.6	Understanding Your Odds	14
3	After an Offer	15
3.1	Compensation Package	15
3.1.1	Equity Grants	15
3.1.2	Bonuses	15
3.1.3	Packages at Different Levels	15
3.2	Negotiation	15
3.2.1	Compensation Expectations	15
3.3	Career Progression	15
4	Where to Start	16
4.1	How Long do I Need for My Job Search?	16
4.2	How Other People Did It	16
4.3	Resources	16
4.3.1	Courses	16
4.3.2	Books and Articles	16
4.3.3	Other Resources	16
4.4	Do's and Don'ts for ML Interviews	16
4.4.1	Do's	16
4.4.2	Don'ts	16
II	Questions	17
5	Math	19
	Notation	19
5.1	Algebra and Calculus	19
5.1.1	Vectors	19
5.1.2	Matricies	19
5.1.3	Dimensionality Reduction	19
5.1.4	Calculus and Convex Optimization	19
5.2	Probability and Statistics	19
5.2.1	Probability	19
5.2.1.1	Basic Concepts to Review	19
5.2.1.2	Questions	19

5.2.2	Statistics	19
6	Computer Science	20
6.1	Algorithms	20
6.2	Complexity and Numerical Analysis	20
6.3	Data	20
6.3.1	Data Structures	20
7	Machine Learning Workflows	21
7.1	Basics	21
7.2	Sampling and Creating Training Data	21
7.3	Objective Functions, Metrics, and Evaluation	21
8	Machine Learning Algorithms	22
8.1	Classical Machine Learning	22
8.1.1	Basic Algorithms	22
8.1.2	Questions	22
8.2	Deep Learning Architectures and Applications	22
8.2.1	Natural Language Processing	22
8.2.2	Computer Vision	22
8.2.3	Reinforcement Learning	22
8.2.4	Other	22
8.3	Training Neural Networks	22
	References	23

Part I

Overview

This part of the book provides an overview of the machine learning interview process, what roles are available, the skills required for each role, the questions that are asked, and how to prepare for them.

Chapter 1

Machine Learning Jobs

1.1 Different Machine Learning Roles

- Applications Engineer
- Applied Research Scientist
- Data Scientist
- Developer Advocate
- Framework Engineer
- ML/AI Infrastructure Engineer
- ML/AI Platform Engineer
- Machine Learning Engineer
- Research Engineer
- Research Scientist
- Solution Architect

1.1.1 Working in Research vs Production

Research and *production* are used instead of *academia* vs *industry*, as most ML research today is carried out by large corporations.

- First thing to figure out is if you want to work in research or production
 - different job descriptions, requirements, hiring processes, and compensation
- Goals of research
 - find answers to fundamental questions
 - expand the body of theoretical knowledge
 - validating a hypothesis without worrying about practicality of results
- Goals of production
 - create or enhance a product
 - product can be goods, services, processes, or business insights
- Research project doesn't need users but product does
- Product has more requirements than research
 - inference latency

- interpretability (users and developers)
- fairness
- adaptability
- production team ensures these other requirements
- Above definitions are loose
- See CS 329S, Lecture 1 for more differences^[2]

1.1.2 Research

- Research community has taken “bigger = better” approach, new models need
 - massive amounts of data
 - tens of millions of dollars in compute
 - unaffordable for most institutions wishing to pursue pure research
- Handful of non-academic ML research labs
 - Google Brain, DeepMind
 - Microsoft
 - Facebook
 - Tencent
- Alphabet accounts for 10+% of all NeurIPS papers
- Not all industry labs publish papers
 - Apple and Tesla notoriously secretive
 - Only publish portions of their research
- Consider publishing policy before joining industry lab
 - Secretive labs may not let you explain to others what you’ve been doing

1.1.2.1 Research vs Applied Research

- Applied research roles
 - somewhere between research and production
 - much closer to research
 - finding solutions to practical problems
 - doesn’t involve implementation in production environments
- Applied researchers are researchers
 - come up with, test, and validate hypotheses
 - need to understand practical problems (subject matter expertise)
- Example research project
 - Develop unsupervised transfer learning method for computer vision
 - Experiment on standard academic dataset
- Example applied research project
 - Apply the new method on real-world problems in a specific industry, e.g. healthcare
 - Need expertise in healthcare and machine learning

1.1.2.2 Research Scientist vs Research Engineer

- Research Scientist
 - come up with novel ideas
 - typically requires PhD and first-authorship at top conferences
 - should be adept engineers themselves
 - may act as an advisor to research engineers
- Research Engineer
 - rare role often seen in major research labs in industry
 - use engineering skills to set up and run experiments for novel ideas
 - doesn't require PhD or publications, but is helpful
- Common to see RS and RE as equal contributors on papers
- Job titles are byproduct of beurocracy
 - RS are supposed to have more academic clout
 - RS paid better than RE
- Startups more generous with job titles to attract talent

1.1.3 Production

ML is finding increasing usage in almost every industry; growing need for bringing ML models to production. This section covers the production ML cycle, required skills, and role distinctions.

1.1.3.1 Production Cycle

1. Project Scoping
 - goals and objectives
 - constraints
 - evaluation criteria
 - stakeholders identified and involved
 - resources estimated and allocated
 - skills
 - product management
 - subject matter expertise
 - some ML knowledge to know what it can and can't solve
2. Data Management
 - data sources, formats, processing, control, and storage
 - data used and generated by ML systems is large and diverse
 - needs scalable infrastructure to process and access it
 - skills
 - databases/query engines for storage/retrieval/processing
 - systems engineering for distributed systems to process lots of data
 - minimal ML knowledge to optimize organization of data for ML

3. ML Model Development

- create training sets from raw data
- data labelling
- feature generation
- model training, optimization, and evaluation
- skills
 - most ML knowledge
 - statistics and probability
 - programming skills for feature engineering and model development

4. Deployment

- make model accessible to users after development
- data-, memory-, and compute-intensive
- skills
 - infrastructure for running ML application
 - may need ML to compress ML models and optimize latency

5. Monitoring and Maintenance

- monitor performance degradation and drift
- maintain/update model to be adaptive in changing environments
- generating and tracking lots of system-generated data (logs)
- skills
 - monitoring and maintenance infrastructure
 - data pipelines

6. Business Analysis

- evaluate model performance against business goals
- analyze model performance to generate business insights
- used to eliminate unproductive projects or scope new ones
- skills
 - ML knowledge to interpret outputs and behavior
 - in-depth statistics to extract insights from data
 - subject matter expertise to map insights to the practical problem

Skills

- **Systems:** systems engineering, building distributed systems, container deployment
- **Databases:** data management, storage, processing, databases, query engines; closely related to **Systems** since distributed systems may be needed for processing lots of data
- **ML::** linear algebra, ML algorithms, etc.
- **Algorithms:** algorithmic programming
- **Statistics:** probability, statistics
- **SME:** subject matter expertise
- **Production:** product management

Most successful approach to ML production is **iterative and incremental development** — lots of back and forth among various steps

Common workflow for building model to predict if an ad should be shown during a search query:

1. Choose a metric to optimize
 - might want to optimize for impressions (number of times ad is shown)
2. Collect data and obtain labels
3. Engineer features
4. Train models
5. Error analysis
 - i. realize errors are caused by wrong labels
 - ii. relabel data
6. Train model again
7. Error analysis
 - i. realize model always predicts ad shouldn't be shown because 99.99% of training data is no-show
 - ii. collect more data on ads that should be shown
8. Train model again
9. Error analysis
 - i. realize model performs well on existing test data from two months ago
 - ii. realize model performs poorly on test data from yesterday
 - iii. model has degraded, collect more recent data
10. Train model again
11. Deploy model
12. Model is performing well
 - i. business people ask why revenue is decreasing
 - ii. realize ads are shown but few people click
 - iii. decide to change model to optimize for clickthrough instead
13. Start over

Many people work on an ML project in production (ML Engineers, Data Scientists, DevOps Engineers, Subject Matter Experts) all from various backgrounds with different languages and tools. They should all be able to work on the system productively. Cross-functional communication and collaboration is essential.

Understanding the production cycle is imperative for candidates:

- idea of what work is needed to bring a model to production
- helps avoid ML projects that are bound to fail

1.1.3.2 ML Engineer vs Software Engineer

- ML engineering considered a subset of software engineering
- Hiring process often spun out of SWE hiring process
 - may swap out a few SWE questions for ML questions

- may add an interview just focused on ML in addition to SWE interviews
- MLE candidates expected to have programming and software engineering skills
- Many traditional SWE tools used for ML model development and deployment
- Companies used to expect stellar SWE and ML researchers but have since relaxed
- Most prefer hiring great engineers but don't know much ML
 - easier for engineers to pick up ML than ML experts picking up good engineering practices

1.1.3.3 ML Engineer vs Data Scientist

- MLEs and Data Scientists spend time wrangling and understanding data
- Reasons for overlap between roles
 - machine learning is part of data science
 - companies want to start with small ML projects before building a full-fledged team
 - data science tasks (including demand forecasting) can be done using ML models
- Goal of Data Science is to generate business insights
 - Data Scientists tend to be better statisticians
- Goal of ML Engineering is to turn data into products
 - MLEs tend to be better engineers
- Many Data Scientists can do their jobs without ML

1.1.3.4 Other Technical Roles in ML Production

Many technical roles in the ML ecosystem don't require any ML knowledge, such as tooling and infrastructure development.

ML Infrastructure Engineer / ML Platform Engineer

- ML relies on infrastructures that scale
- Companies with mature ML often have dedicated infrastructure teams
- skills
 - parallelism
 - distributed computing
 - low-level optimization
- hard skills to learn, take time to master
 - companies prefer hiring skilled systems engineers, then training them in ML

ML Accelerator / Hardware Engineer

- hardware is a major bottleneck
- ML algorithms are constrained by slow compute, limited memory, operating cost, on-device power, etc.
- explosion of dedicated ML hardware
- hardware companies need people with ML expertise to guide hardware development
- also looking to use ML algorithms to improve chip design

ML Solutions Architect

- often found in professional services companies
- each company has unique use-cases and requirements
- involves working with existing or potential customers
 - figure out if you can help with their use-case
 - design solution for their use-case

Developer Advocate

- first users of ML products
- writing tutorials, giving talks
- community engagement
- hard to fill roles
 - difficult to find great engineers with matching communication skills

1.1.3.5 Understanding Roles and Titles

- people on the same team with the same title may do very different things
 - Lyft renamed “Data Analyst” → “Data Scientist” and “Data Scientist” → “Research Scientist”

1.2 Types of Companies

1.2.1 Applications Companies vs Tooling Companies

1.2.2 Enterprise vs Consumer Products

1.2.3 Startups or Big Companies

Chapter 2

Machine Learning Interview Process

2.1 Understanding the Interviewers' Mindset

2.1.1 What Companies Want From Candidates

2.1.1.1 Technical Skills

2.1.1.2 Non-technical Skills

2.1.1.3 What Exactly is Culture Fit?

2.1.1.4 Junior vs Senior Roles

2.1.1.5 Do I need a PhD to Work in ML?

2.1.2 How Companies Source Candidates

2.1.3 What Signals Companies Look for in Candidates

2.2 Interview Pipeline

2.2.1 Common Interview Formats

2.2.2 Alternative Interview Formats

2.2.3 Interviews at Big Companies vs at Small Companies

2.2.4 Interviews for Internships vs for Full-time Positions

2.3 Types of Questions

2.3.1 Behavioral Questions

2.3.1.1 Background and Resume

2.3.1.2 Interests

2.3.1.3 Communication

2.3.1.4 Personality

Chapter 3

After an Offer

3.1 Compensation Package

3.1.1 Equity Grants

3.1.2 Bonuses

3.1.3 Packages at Different Levels

3.2 Negotiation

3.2.1 Compensation Expectations

3.3 Career Progression

Chapter 4

Where to Start

4.1 How Long do I Need for My Job Search?

4.2 How Other People Did It

4.3 Resources

4.3.1 Courses

4.3.2 Books and Articles

4.3.3 Other Resources

4.4 Do's and Don'ts for ML Interviews

4.4.1 Do's

4.4.2 Don'ts

Part II

Questions

This part of the book contains over 200 knowledge questions of varying difficulty. They cover important concepts and common misconceptions in machine learning. More senior roles should expect harder questions. Several questions are technically incorrect or ambiguous to see if candidates will correct the interviewer, point out edge cases, or ask for clarification.

There are accompanying hints for these questions that should clarify ambiguity or technical incorrectness. The first draft of the book has answers for about 10% of questions, though more can be found as pull-requests in the [book repository](#).}

Chapter 5

Math

Notation

- All vectors are column vectors of dimension $n \times 1$, with n being the number of elements
- The i^{th} element in the dataset is denoted $\mathbf{x}^{(i)}$
- The i^{th} row of an $n \times m$ matrix is \mathbf{x}_i

5.1 Algebra and Calculus

5.1.1 Vectors

5.1.2 Matrices

5.1.3 Dimensionality Reduction

5.1.4 Calculus and Convex Optimization

5.2 Probability and Statistics

5.2.1 Probability

5.2.1.1 Basic Concepts to Review

5.2.1.2 Questions

5.2.2 Statistics

Chapter 6

Computer Science

6.1 Algorithms

6.2 Complexity and Numerical Analysis

6.3 Data

6.3.1 Data Structures

Chapter 7

Machine Learning Workflows

7.1 Basics

7.2 Sampling and Creating Training Data

7.3 Objective Functions, Metrics, and Evaluation

Chapter 8

Machine Learning Algorithms

8.1 Classical Machine Learning

8.1.1 Basic Algorithms

8.1.2 Questions

8.2 Deep Learning Architectures and Applications

8.2.1 Natural Language Processing

8.2.2 Computer Vision

8.2.3 Reinforcement Learning

8.2.4 Other

8.3 Training Neural Networks

References

- [1] C. Huyen, *Machine Learning Interviews*. 2021. Available: <https://huyenchip.com/ml-interviews-book/>
- [2] C. Huyen, *CS 329S: Machine learning systems design*. 2021. Available: <https://cs329s.stanford.edu>