**MACHINE LEARNING INTERVIEW QUESTIONS: PROGRAMMING**

These machine learning interview questions test your knowledge of programming principles you need to implement machine learning principles in practice. Machine learning interview questions tend to be technical questions that test your logic and programming skills: this section focuses more on the latter.

Q26: How do you handle missing or corrupted data in a dataset?

Answer: You could find missing/corrupted data in a dataset and either drop those rows or columns, or decide to replace them with another value.

In Pandas, there are two very useful methods: isnull() and dropna() that will help you find columns of data with missing or corrupted data and drop those values. If you want to fill the invalid values with a placeholder value (for example, 0), you could use the fillna() method.

More reading: Handling missing data (O’Reilly)

Q27: Do you have experience with Spark or big data tools for machine learning?

Answer: You’ll want to get familiar with the meaning of big data for different companies and the different tools they’ll want. Spark is the big data tool most in demand now, able to handle immense datasets with speed. Be honest if you don’t have experience with the tools demanded, but also take a look at job descriptions and see what tools pop up: you’ll want to invest in familiarizing yourself with them.

More reading: 50 Top Open Source Tools for Big Data (Datamation)

Q28: Pick an algorithm. Write the pseudo-code for a parallel implementation.

Answer: This kind of question demonstrates your ability to think in parallelism and how you could handle concurrency in programming implementations dealing with big data. Take a look at pseudocode frameworks such as Peril-L and visualization tools such as Web Sequence Diagrams to help you demonstrate your ability to write code that reflects parallelism.

More reading: Writing pseudocode for parallel programming (Stack Overflow)

Q29: What are some differences between a linked list and an array?

Answer: An array is an ordered collection of objects. A linked list is a series of objects with pointers that direct how to process them sequentially. An array assumes that every element has the same size, unlike the linked list. A linked list can more easily grow organically: an array has to be pre-defined or re-defined for organic growth. Shuffling a linked list involves changing which points direct where—meanwhile, shuffling an array is more complex and takes more memory.

More reading: Array versus linked list (Stack Overflow)

Q30: Describe a hash table.

Answer: A hash table is a data structure that produces an associative array. A key is mapped to certain values through the use of a hash function. They are often used for tasks such as database indexing.

More reading: Hash table (Wikipedia)

Q31: Which data visualization libraries do you use? What are your thoughts on the best data visualization tools?

Answer: What’s important here is to define your views on how to properly visualize data and your personal preferences when it comes to tools. Popular tools include R’s ggplot, Python’s seaborn and matplotlib, and tools such as Plot.ly and Tableau.

More reading: 31 Free Data Visualization Tools (Springboard)

Related: 20 Python Interview Questions

Q32: Given two strings, A and B, of the same length n, find whether it is possible to cut both strings at a common point such that the first part of A and the second part of B form a palindrome.

Answer: You’ll often get standard algorithms and data structures questions as part of your interview process as a machine learning engineer that might feel akin to a software engineering interview. In this case, this comes from Google’s interview process. There are multiple ways to check for palindromes—one way of doing so if you’re using a programming language such as Python is to reverse the string and check to see if it still equals the original string, for example. The thing to look out for here is the category of questions you can expect, which will be akin to software engineering questions that drill down to your knowledge of algorithms and data structures. Make sure that you’re totally comfortable with the language of your choice to express that logic.

More reading: Glassdoor machine learning interview questions

Q33: How are primary and foreign keys related in SQL?

Answer: Most machine learning engineers are going to have to be conversant with a lot of different data formats. SQL is still one of the key ones used. Your ability to understand how to manipulate SQL databases will be something you’ll most likely need to demonstrate. In this example, you can talk about how foreign keys allow you to match up and join tables together on the primary key of the corresponding table—but just as useful is to talk through how you would think about setting up SQL tables and querying them.

More reading: What is the difference between a primary and foreign key in SQL?

Q34: How does XML and CSVs compare in terms of size?

Answer: In practice, XML is much more verbose than CSVs are and takes up a lot more space. CSVs use some separators to categorize and organize data into neat columns. XML uses tags to delineate a tree-like structure for key-value pairs. You’ll often get XML back as a way to semi-structure data from APIs or HTTP responses. In practice, you’ll want to ingest XML data and try to process it into a usable CSV. This sort of question tests your familiarity with data wrangling sometimes messy data formats.

More reading: How Can XML Be Used?

Q35: What are the data types supported by JSON?

Answer: This tests your knowledge of JSON, another popular file format that wraps with JavaScript. There are six basic JSON datatypes you can manipulate: strings, numbers, objects, arrays, booleans, and null values.

More reading: JSON datatypes

Q36: How would you build a data pipeline?

Answer: Data pipelines are the bread and butter of machine learning engineers, who take data science models and find ways to automate and scale them. Make sure you’re familiar with the tools to build data pipelines (such as Apache Airflow) and the platforms where you can host models and pipelines (such as Google Cloud or AWS or Azure). Explain the steps required in a functioning data pipeline and talk through your actual experience building and scaling them in production.

More reading: 10 Minutes to Building A Machine Learning Pipeline With Apache Airflow

Machine Learning Interview Questions: Company/Industry Specific

These machine learning interview questions deal with how to implement your general machine learning knowledge to a specific company’s requirements. You’ll be asked to create case studies and extend your knowledge of the company and industry you’re applying for with your machine learning skills.

Q37: What do you think is the most valuable data in our business?

Answer: This question or questions like it really try to test you on two dimensions. The first is your knowledge of the business and the industry itself, as well as your understanding of the business model. The second is whether you can pick how correlated data is to business outcomes in general, and then how you apply that thinking to your context about the company. You’ll want to research the business model and ask good questions to your recruiter—and start thinking about what business problems they probably want to solve most with their data.

More reading: Three Recommendations For Making The Most Of Valuable Data

Q38: How would you implement a recommendation system for our company’s users?

Answer: A lot of machine learning interview questions of this type will involve the implementation of machine learning models to a company’s problems. You’ll have to research the company and its industry in-depth, especially the revenue drivers the company has, and the types of users the company takes on in the context of the industry it’s in.

More reading: How to Implement A Recommendation System? (Stack Overflow)

Q39: How can we use your machine learning skills to generate revenue?

Answer: This is a tricky question. The ideal answer would demonstrate knowledge of what drives the business and how your skills could relate. For example, if you were interviewing for music-streaming startup Spotify, you could remark that your skills at developing a better recommendation model would increase user retention, which would then increase revenue in the long run.

The startup metrics Slideshare linked above will help you understand exactly what performance indicators are important for startups and tech companies as they think about revenue and growth.

More reading: Startup Metrics for Startups (500 Startups)

Q40: What do you think of our current data process?

machine learning interview questions

Answer: This kind of question requires you to listen carefully and impart feedback in a manner that is constructive and insightful. Your interviewer is trying to gauge if you’d be a valuable member of their team and whether you grasp the nuances of why certain things are set the way they are in the company’s data process based on company or industry-specific conditions. They’re trying to see if you can be an intellectual peer. Act accordingly.

More reading: The Data Science Process Email Course (Springboard)

Machine Learning Interview Questions: General Machine Learning Interest

This series of machine learning interview questions attempts to gauge your passion and interest in machine learning. The right answers will serve as a testament to your commitment to being a lifelong learner in machine learning.

Q41: What are the last machine learning papers you’ve read?

Answer: Keeping up with the latest scientific literature on machine learning is a must if you want to demonstrate an interest in a machine learning position. This overview of deep learning in Nature by the scions of deep learning themselves (from Hinton to Bengio to LeCun) can be a good reference paper and an overview of what’s happening in deep learning — and the kind of paper you might want to cite.

More reading: What are some of the best research papers/books for machine learning?

Q42: Do you have research experience in machine learning?

Answer: Related to the last point, most organizations hiring for machine learning positions will look for your formal experience in the field. Research papers, co-authored or supervised by leaders in the field, can make the difference between you being hired and not. Make sure you have a summary of your research experience and papers ready—and an explanation for your background and lack of formal research experience if you don’t.

Q43: What are your favorite use cases of machine learning models?

Answer: The Quora thread below contains some examples, such as decision trees that categorize people into different tiers of intelligence based on IQ scores. Make sure that you have a few examples in mind and describe what resonated with you. It’s important that you demonstrate an interest in how machine learning is implemented.

More reading: What are the typical use cases for different machine learning algorithms? (Quora)

Q44: How would you approach the “Netflix Prize” competition?

Answer: The Netflix Prize was a famed competition where Netflix offered $1,000,000 for a better collaborative filtering algorithm. The team that won called BellKor had a 10% improvement and used an ensemble of different methods to win. Some familiarity with the case and its solution will help demonstrate you’ve paid attention to machine learning for a while.

More reading: Netflix Prize (Wikipedia)

Q45: Where do you usually source datasets?

Answer: Machine learning interview questions like these try to get at the heart of your machine learning interest. Somebody who is truly passionate about machine learning will have gone off and done side projects on their own, and have a good idea of what great datasets are out there. If you’re missing any, check out Quandl for economic and financial data, and Kaggle’s Datasets collection for another great list.

More reading: 19 Free Public Data Sets For Your First Data Science Project (Springboard)

Q46: How do you think Google is training data for self-driving cars?

Answer: Machine learning interview questions like this one really test your knowledge of different machine learning methods, and your inventiveness if you don’t know the answer. Google is currently using recaptcha to source labeled data on storefronts and traffic signs. They are also building on training data collected by Sebastian Thrun at GoogleX—some of which was obtained by his grad students driving buggies on desert dunes!

More reading: Waymo Tech

Q47: How would you simulate the approach AlphaGo took to beat Lee Sedol at Go?

Answer: AlphaGo beating Lee Sedol, the best human player at Go, in a best-of-five series was a truly seminal event in the history of machine learning and deep learning. The Nature paper above describes how this was accomplished with “Monte-Carlo tree search with deep neural networks that have been trained by supervised learning, from human expert games, and by reinforcement learning from games of self-play.”

More reading: Mastering the game of Go with deep neural networks and tree search (Nature)

Q48: What are your thoughts on GPT-3 and OpenAI’s model?

Answer: GPT-3 is a new language generation model developed by OpenAI. It was marked as exciting because with very little change in architecture, and a ton more data, GPT-3 could generate what seemed to be human-like conversational pieces, up to and including novel-size works and the ability to create code from natural language. There are many perspectives on GPT-3 throughout the Internet — if it comes up in an interview setting, be prepared to address this topic (and trending topics like it) intelligently to demonstrate that you follow the latest advances in machine learning.

More reading: Language Models are Few-Shot Learners

Q49: What models do you train for fun, and what GPU/hardware do you use?

Answer: This question tests whether you’ve worked on machine learning projects outside of a corporate role and whether you understand the basics of how to resource projects and allocate GPU-time efficiently. Expect questions like this to come from hiring managers that are interested in getting a greater sense behind your portfolio, and what you’ve done independently.

More reading: Where to get free GPU cloud hours for machine learning

Q50: What are some of your favorite APIs to explore?

Answer: If you’ve worked with external data sources, it’s likely you’ll have a few favorite APIs that you’ve gone through. You can be thoughtful here about the kinds of experiments and pipelines you’ve run in the past, along with how you think about the APIs you’ve used before.

More reading: Awesome APIs

Q51: How do you think quantum computing will affect machine learning?

Answer: With the recent announcement of more breakthroughs in quantum computing, the question of how this new format and way of thinking through hardware serves as a useful proxy to explain classical computing and machine learning, and some of the hardware nuances that might make some algorithms much easier to do on a quantum machine. Demonstrating some knowledge in this area helps show that you’re interested in machine learning at a much higher level than just implementation details.

More reading: Quantum Machine Learning