Data Science Talk

IIT Bombay - 9/12/2020

Akshit Arora, Deepshikha Kumari

Portfolio:

1. Kaggle: Your Machine Learning and Data Science Community - can be used to demonstrate skills.
2. Explore GitHub - upload projects here, with up-to-date README.md file and clear instructions to reproduce the results.
3. Have a blog where you write about your experiences in ML/DL.

Blogs/ Talks:

1. Panel: Advancing Your Data Science Career During the Pandemic - #380 (This Week in Machine Learning Podcast)
2. Hackerrank, CodeForces, Coding Ninjas are also some of the top platforms you can head over to.
3. If you're an absolute beginner and want to learn Python, start here https://www.learnpython.org/.
4. Statistics
    a. Book on Practical Statistics — This will teach you statistics from a Data Science standpoint. You should read at least the first three chapters of this book.
    b. Statistics and Probability | Khan Academy — This course will prepare you well for all the statistics and probability-related questions during the interview. A free course with a good compilation of video lectures and practice problems.
    c. Naked Statistics — For people who dread mathematics and prefer to understand practical examples, this is a fantastic book explaining how statistics are applied in real-life scenarios.
5. Data Wrangling using SQL/NoSQL/Pandas:
    a. Kaggle Learn — You should check out Kaggle's series of courses here that cover SQL, Pandas, data visualizations, and much more.
    b. DataCamp Data Analyst Python Track — The course would also get you up to speed with skills that you'd need as an analyst, especially when it comes to data manipulation and visualizations. They provide interactive, hands-on exercises on each of the tools mentioned above.
    c. Data Analysis with Python — The course covers all the essential skills for a Data Analyst.

    d. Practice Questions — Once you have a good understanding, read through a few example questions on blogs and career websites like springboard. (respond to this post if you'd like me to prepare a list of questions)
6. Machine Learning Algorithms
    a. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow, 2nd Edition
7. Machine Learning projects:
    a. Practical Deep Learning for Cloud, Mobile, and Edge: Real-World AI & Computer-Vision Projects Using Python, Keras & TensorFlow (GitHub)
8. Organize / Volunteer at community events around machine learning / deep learning.

Data Science Roles:

What all is available in the world of Data Science in a typical company. And what interests you more?

1. Data Analysts:

Example Requirements:

- Develop analytics solutions to support business analytics such as marketing analytics, customer segmentation, acquisition, business operations, retention, and forecasting.
- Develop statistical testing programs to measure and improve on-going performance.
- Apply appropriate techniques, such as exploratory data analysis, regression analysis, bootstrapping, decision trees, cluster analysis, choice modeling, survival analysis, etc.

You should be focussing on preparing the following topics:

- Data Cleaning/Wrangling — Transforming data and deriving insight from it.
    - Large data processing -- data analyst spends more time understanding the data.
- Basics of Statistics — Fundamentals of probability, statistics, and linear algebra.
- Python/SQL refresher
- Google Analytics (based on the Job description)
- Excel and Tableau Refresher for those interested in Business Intelligence Analysts and Marketing Analysts
- More machine learning than deep learning.
- Case study interview practice

2. Data Engineers

More technical and doesn't need as much math and stats. Having a solid grip over programming and understanding of data management tools is essential. These Engineers serve as DevOps for Data Science. You should be focussing on preparing the following topics:

- Competent Programming/Development skills, knowledge of web frameworks, REST APIs, etc.

- Big Data tools — spark, Hadoop, airflow(DevOps). Learn to develop distributed systems.
- Developing ETL pipelines.
- Cloud Service Providers (Google Cloud or AWS based on the JD)

3. ML/AI Engineers

Example Requirements:

- End-to-end/full stack Machine Learning (ML) Algorithm and Model development.
- Design, code, test, debug, validate, and productionalise models; document design decisions and develop modular software components for model inferences; monitor system performance metrics, and identify potential risks/issues
- Select appropriate data-sets and data representation methods, actively managing data risks and issues
- Perform statistical analysis and fine-tuning models and create globally generalizable models performing well on real-time inputs
- Work with product owners to understand desired application capabilities and testing scenarios, execute against strategic technical management and delivery roadmaps
- Operate in an Agile/Scrum environment to deliver high-quality models and products

AI Engineers are expected to have a deep understanding of ML algorithms; they should be aware of the mathematics that drives those algorithms to optimized states.

Following are the essential concepts you should look to master:

- Good Programming skills, for obvious reasons.
- Study various loss functions, cost functions, training algorithms, regularization methods, optimizers for neural networks, etc. The hands-on book would come in very handy to prepare these.
- Good grip over the mathematical and statistical foundation of algorithms — linear algebra, multivariate calculus, stochastic analysis at a minimum.
- Working with massive datasets, best practices to implement batch learning, and online learning.
- Familiarity with Data Engineering tools — You may be required to deploy the models yourself or help the engineering team integrate your deep models to work as expected on production servers.

4. Data Scientists

The majority of companies look for generalists to join their team of data scientists. The position expects you to design and execute A/B testing experiments, perform statistical analysis on samples of data, sometimes refactor production code and visualize data.

Example Requirements:

- Develop highly scalable classifiers and tools leveraging machine learning and Natural Language Processing or Computer vision
- Work as part of the product team to implement algorithms that scale on billions of text messages.
- Be responsible for measuring and optimizing the quality of your algorithms and Models.

Thus the topics you should prepare are:

- Good programming skills (Python, R, Matlab).
- Deep understanding of mathematical and statistical concepts for designing and conducting research experiments and evaluating their significance of the observations.
- Should have studied classification and regression problems, learn to deal with supervised and unsupervised learning.
- Proven ability to solve complex problems like image classification, speech recognition, natural language processing. Learn to use pre-trained models; experience with libraries like TensorFlow, PyTorch.
- Domain Expertise — This depends on your area of research and the industry that the company belongs to. For example, a research scientist at a hedge fund needs to have a deep understanding of finance and economics.
- Familiarity with Data Engineering tools — They should also have a brief understanding of how data infrastructure is set up and how their code goes to production servers.

Data Science at NVIDIA

Solutions Architect Role. Basic responsibilities include:

1. Develop and demonstrate solutions based on NVIDIA's state-of-the-art ML/DL, data science software, and hardware technologies to customers
2. Perform in-depth analysis and optimization to ensure the best performance on GPU architecture systems.
3. Applying deep learning solutions to areas such as object detection, segmentation, video understanding, sequence prediction, adaptive computing, memory networks, reduced precision training and inference, graph compilers, reinforcement learning, search, distributed and federated training, and more.
4. Collaborate with key industry partner/customer developers to provide ML solutions applied to their products and technologies.
5. Partner with Engineering, Product, and Sales teams to secure design wins at customers.
6. Work closely with customer's data science, ML/DL developers, and IT teams.