


Machine Learning in Action

 [appliedmachinelearning.blog/2018/04/13/my-data-science-machine-learning-job-interview-experience-list-of-ds-ml-dl-](https://appliedmachinelearning.blog/2018/04/13/my-data-science-machine-learning-job-interview-experience-list-of-ds-ml-dl-questions/)

Abhijeet Kumar

April 13, 2018



Beginning this year, I had been taking interviews for a job change in Data Science (DS)/Machine Learning (ML)/Deep Learning (DL) field in India. In the process of 30-40 days of job hunt, I took interviews in about 8-10 companies inclusive of start-ups, services based and product based in nature. I believe, It would help job seekers; hence I decided to blog post my experience and what I learned along the way. The blog-post lists 100 of data science interview questions.

A bit of introduction first, I have 4+ years of experience in machine learning and its applications in field of speech analytics, text analytics and image analytics. Somehow, in general I felt that most of the job positions in this area revolve around text analytics (Natural language processing) followed by image analytics (Computer vision). There are very few companies pursuing speech/audio analytics. Having 5 years of experience, I was looking for a mid-level senior position where I can lead a team of DS/ML enthusiasts to do some interesting work/projects. I will be listing down the kind of questions that were asked in various technical interviews of recruitment process.

Company 1 (20-25 minutes) – Global service based company

1. As in your CV, what have you done in building document exploration system ? I have implemented document clustering through topic modeling with Latent Dirichlet Allocation (LDA) technique. Find my posts [here](#), [here](#) and [here](#) on this topic.
2. Let's say, you are given a scenario where you have terabytes of data files consisting of pdfs, text files, images, scanned pdfs etc. What approach will you take in understanding or classifying them ?
3. How will you read the content of scanned pdfs or written documents in image formats ? Well, here is a [blog-post](#) on solution to this task.
4. Why is naive bayes called "naive" ?
5. Tell me about naive bayes classifier ?

6. What is deep learning ? Difference between machine learning & Deep learning ?

Experience:

There were few more questions like this but it went terrible for me. I really could not get what his expectation was as he was not going technical at all where as I was trying to dig deep into technical stuff. I tried discussing about training a tesseract or language models but he does not seem to be convinced. May be he wanted to hear something out of box or may be just good explanations or may be some better approach. I did not get the difference between getting interviewed as fresher and experienced professional as I was taking it for the first time in last 5 years.

Company 2 (40-45 minutes) – Global service based company

1. How can you cluster documents in unsupervised way ?
2. How do you find the similar documents related to some query sentence/search ?
3. Explain TF-IDF ?
4. So As per my experience, Tf-Idf fails in document classification/clustering ? How can you improve further ?
5. What is LSTM neural network ? Explain me !! How does it work ?
6. What are word2vec vectors?
7. Explain mutable and immutable objects in python ?
8. Tell about the data structures you have used in python ?

Experience:

Though there was lot of discussion around document similarity problem, I was through the interview process. Again, there was not much interest in going to technical depth. Mostly the company had short automation projects in text analytics. I was offered the ML architect job position.

Company 3 – (40 minutes) Global product & service based

1. How do you handle multi-class classification with unbalanced dataset ?
2. How did you perform language identification from text sentence ? (As in my CV). Here is a blog-post that demonstrates the solution.
3. How does you represent the symbolic chinese or japanese alphabets here ?
4. How can I design a chatbot ? (I had little idea but I tried answering it with intent and response tf-idf based similarity)
5. Can I develop a chatbot with RNN providing a intent and response pair in input ?
6. Suppose I developed a chatbot with RNN/LSTMs on Reddit dataset. It gives me 10 probable responses ? How can I choose the best reply Or how can I eliminate others replies ?
7. How does SVM learns non-linear boundaries ? Explain.

Experience:

There were few more questions which I do not remember now. This was the first technical interview where I had to go on explaining technical details which made me happy. I was offered job position in the company.

Company 4 – (50 minutes) – 1 year old Heath care start-up

1. What is precision and recall ? Which one of this do you think is important in medical diagnosis ?
2. Define precision and recall ?
3. How do you draw ROC curve ? What does area under ROC curve signify ?
4. How will you draw ROC for multi class classification problem ?
5. Tell me any other metric to measure multi-class classification result ?
6. What is sensitivity and specificity ?
7. What is random about Random Forest ?
8. How do you perform text classification ?
9. How can you make sure to learn a context !! Well its not possible with TF-IDF ? (I told him about taking n-grams say $n = 1, 2, 3, 4$ and concatenating tf-idf of them to make a long count vector ?
10. Okay that is the baseline people start with ? What can you do more with machine learning ? (I tried suggesting LSTM with word2vec or 1D-CNN with word2vec for classification but he wanted improvements in machine learning based methods :-)
11. How does neural networks learns non-linear shapes when it is made of linear nodes ? What makes it learn non-linear boundaries ?

Experience:

There were few more good questions which I do not remember now. Though the interview discussion was fantastic, in some of the questions we were not on same platform. Also during the interview I found out that being a startup there were only 2-3 people working in ML/DL/DS and showed the concern to the interviewer. I was not selected for the position here.

Company 5 – (50-55 minutes) – Amazon Inc.

1. What are the parameters in training a decision tree ?
2. Explain the criteria for splitting at a node in decision trees ?
3. Tell us the formula of Gini index criteria?
4. What is the formula for Entropy criteria?
5. How is it decided that on which features it has to split ?
6. How do you calculate information gain mathematically ? Are you sure of formula!!
7. What is the advantage with random forest ?
8. Tell me about boosting algorithms ?
9. Okay !! So how does gradient boosting works ?
10. Do you know about Adaboost algorithm ? How and why does it work ?
11. What are the kernels used in SVM ? What is the optimization technique of SVM ?

12. How does SVM learn the hyperplane ? Talk more about mathematical details ?
13. Talking about unsupervised learning ? What are the algorithms ?
14. How do you decide K in K-Means clustering algorithm ?
15. Tell me at least 3 ways of deciding K in clustering ?
16. What other clustering algorithms do you know ?
17. Okay !! Can you tell DB-SCAN algorithm ?
18. How does HAC (Hierarchical Agglomerative clustering) work ?
19. Explain PCA ? Tell me the mathematical steps to implement PCA ?
20. What is disadvantage of using PCA ?
21. How does CNN work ? Explain the implementation details ?
22. Explain the back propagation steps of Convolution Neural Network ?
23. How do you deploy Machine Learning models ?
24. Lot of times, we may have to write ML models from scratch in C++ ? Will you be able to do that ?

Experience:

This was Amazon interview for level 6 position. All I can say that it was mainly focused on algorithms and their mathematics behind them. Unfortunately, I was taking all interviews extempore and was not in a position to discuss all the mathematics in detailed way. Though I really enjoyed the interview discussing at least all the methodologies and mathematical explanations (which I remembered), the interviewer does not find me suitable for Level 6 position. I suppose one can easily crack initial technical round if you brush up the common machine learning algorithms in detail.

Company 6 – (50-55 minutes) – Global service based Company

1. What is the range of sigmoid function ?
2. Name the package of scikit-learn that implements logistic regression ?
3. What is mean and variance of standard normal distribution ?
4. What are the data structures you have used in python ?
5. Text classification method. How will you do it ?
6. Explain Tf-Idf ? What is the drawback of Tf-Idf ? How do you overcome it ?
7. What are bigrams & Tri-grams ? Explain with example of Tf-Idf of bi-grams & trigrams with a text sentence.
8. What is an application of word2vec ? Example.
9. How will you design a neural network ? How about making it very deep ? Very basic questions on neural network.?
10. How does LSTM work ? How can it remember the context ?
11. What is naive bayes classifier ?
12. What is the probability of heads coming 4 times when a coin is tossed 10 number of times ?
13. How do you get an index of an element of a list in python ?
14. How do you merge two data-set with pandas ?
15. From user behavior, you need to model fraudulent activity. How are you going to solve this ?
May be anomaly detection problem or a classification problem !!

16. What will you prefer a decision tree or a random forest ?
17. How is using a logistic regression different from using a random forest ?
18. Will you use decision tree or random forest for a classification problem ? What is advantage of using random forest ?

Experience:

Well I was offered the data scientist job position. In-fact I got a very good feedback. It was a good technical discussion which I enjoyed. Also, you may feel that these questions are basic in ML/DS field. I do not want to demean but somewhere I felt that interviewer may not have worked much in the domain or may not be aware of current stuff going in the field.

Company 7 (25-30 minutes) – Global business process management company

1. Which model would you use in case of unbalanced dataset: Random Forest or Boosting ? Why ?
2. What are the boosting techniques you know ?
3. Which model would you like to choose if you have many classes in a supervised learning problem ? Say 40-50 classes !!
4. How do you perform ensemble technique?
5. How does SVM work ?
6. What is Kernel ? Explain a few.
7. How do you perform non-linear regression ?
8. What are Lasso and Ridge regression ?

Experience:

Frankly speaking, the technical interview was little vague (personal opinion) as I do not find the conversation serious enough. But questions were good. The job position was for a senior position where I would be leading 15-16 people team. It was followed by a client interview and HR interview. I was offered the consultant job position with good compensation.

Company 8 (60 minutes) – 4 years old product & services company

1. As per CV, you have done speaker identification in speech. Which are the approaches you followed ?. Find my blog-post here on this topic.
2. What are MFCCs ?
3. What is Gaussian Mixture model ? How does it perform clustering ?
4. How is Expectation Maximization performed ? Explain both the steps ?
5. How is likelihood calculated in GMM ?
6. How did you perform MAP adaptation for GMM-UBM technique of speaker identification ?
7. Tell me about I-vector technique you implemented ?

8. Okay !! What is factor analysis in this context ?
9. What was the difference between JFA and I-vector ? Why choose I-vectors over JFA
10. Have you implemented PLDA I-vector technique ?
11. Have you read Deep Speaker paper from Baidu ?
12. How do you select between 2 models (Model Selection techniques)?
13. Okay great BIC & AIC !! How does it work mathematically ?
14. Explain the intuition behind BIC or AIC ?
15. How do you handle missing data or NaNs in your MFCC feature vector matrix ?
16. How did you perform language identification ? What were the feature ?
17. How did you model classifiers like speech vs music and speech vs non-speech ?
18. How can deep neural network be applied in these speech analytics applications ?

Experience:

Yeah, You may be thinking that what kind of questions are these. Coincidentally the area of research for both of us was speech analytics (especially speaker identification). So, the whole 1 hour of interview went on discussing the speech analytics domain. Obviously, He was very well convinced with the work and gave positive feedback for the interview. I was offered AI solution architect job position after one more round with co-founders.

Suggestions

I have spoken to almost 25-30 professionals working in this fields during this job change process. Having said that, I would suggest the following tips which may support the job seekers.

- Your CV is very important. Start building your profile impressive with projects, Kaggle competitions, MOOC certifications, research papers (if possible). I got call from amazon without any reference. Your CV can impress HR's and interviewer right away. It is certainly hard to get picked for interview in good companies without reference.
- Your confidence and drive that you carry will do half of the job. It's very important to take interview with confidence and show that you have that drive. Personally I felt it matters in interviews (specially in startups and services based ones).
- While in interview do not rush into answers. Take your time in thinking about questions and then answer suitably. Ask questions if you do not understand the question properly. You may need to calm down in interviews (that's an issue with me sometimes)
- Express yourself while explaining concepts properly. Give examples from projects you may have implemented. You must be thorough with the skills and projects written in CV.
- Mostly, Employer looks for experienced candidates in this field. If you are a beginner then try to build your profile by implementing applications on your own. Your Github account can say a lot. Do participate in lot of Kaggle competitions. Take online MOOC certifications.
- Lastly try to be modest with interviewer and nod to their opinions sometimes otherwise it may land you in rejections. You may find people arrogant sometimes. For an example, R or Python enthusiasts are sentimental about respective languages. It is better not to fall in such debate. Personally I feel these are tools to accomplish your logic and concepts.

Best of Luck.

Hope these data science and machine learning interview questions will help the beginners for their job preparations.

If you liked the post, Kindly share it so that it can reach out to the readers who can actually gain from this. Follow my blog to get updates about upcoming articles on Machine learning or Deep Learning.

I would love to answer your query if any.

Happy interviewing

Advertisements

Report this ad

Report this ad