CODING CLUB
IIT GUWAHATI

DATA SCIENCE INTERN 101

General Advice

- Get your resume scrutinized by seniors
- Don't give cliched answers in interviews
- Be ready with answers to questions which can be asked, for example, questions related to your resume
- In coding tests, if nothing works, try passing partial test cases
- Decent coding skills are must for any kind of internship.
- Be Confident in interviews, give as many mock interviews as possible
- There is no such one-size-fits-all course which you can do for an exhaustive prep, it is best to go over many websites and blogs to learn once you know what topics to keep an eye out for, a good blog to refer is- Towards Data Science
- Try to read blogs daily of different concepts and see kaggle notebooks of top participants of any dataset
- Practice on some datasets on kaggle or take part in competitions (at least 3-4)
- For the ml internships you need to have good knowledge of probability and statistics
- For the traditional ml part you need to know the maths behind all the algorithms (that is a must)
- Basic questions are usually asked(almost none from deep learning).
- Basic ML concepts should be clear.

Topics and resources

1. ***Linear algebra and puzzles***
    a. Linear Algebra Lectures- Gilbert Strang, very good video lecture series to go through if you have the time.
    b. Puzzles in analytics interviews:-- Brainstellar - Puzzles from Quant Interviews

2. ***Probability and Statistics***
    a. Descriptive Statistics topics
        i. Mean, Median, Mode
        ii. Variance and Standard Deviation

b. Probability topics
    i. Bernoulli Trials & Probability Mass Function
    ii. Central Limit Theorem
    iii. Normal Distribution
c. Inferential Statistics topics
    i. Confidence Interval
    ii. Hypothesis Testing
    iii. Correlation
d. Some good books for concepts are :
    - Introduction to Probability, 2nd Edition by Dimitri P. Bertsekas and John N. Tsitsiklis
    - R. V. Hogg, J. W. McKean and A. Craig, Introduction to Mathematical Statistics, 6th Ed, Pearson Education India, 2006
e. Probability - you should know all types of distribution, joint probability, marginal probability etc. Try to solve as many problems as possible on expectation. A good textbook to practice this is - Introduction to Probability, Statistics, and Random Processes by Hossein Pishro-Nik
f. The Stat110 course is amazing for probability (random variables). Statistics 110: Probability
g. fifty_challenging_problems_in__2.pdf
h. Statistics - you should know the central limit theorem, p-statistic, t-statistic, chi and other tests, all concepts about standard deviation, variance, correlation, covariance. Tests such as chi squared test, Shapiro test etc.
   Course : Statistics with R Specialization from Coursera
i. A blog on inferential stats:
   Comprehensive & Practical Inferential Statistics Guide for data science
j. Detailed guide for hypothesis testing:
   Master Hypothesis Testing in Statistics

3. ***Machine Learning***
a. Reddit - /r/MachineLearning , /r/datascience , /r/learnmachinelearning
b. Youtube playlist for an overview : Machine Learning Recipes with Josh Gordon
c. Crash course for basic ML : Google Developers Machine Learning Crash Course
d. Traditional ml topics to know - data preprocessing technique which includes missing values imputation, outliers detection, pandas, eda, metrics etc.

e. Basic algorithms such as linear regression, logistic regression, dimension reduction algorithms etc. Go through the CS229 course. It will give you mathematical understanding about basic concepts(Maths behind linear/logistic regression) Stanford CS229: Machine Learning | Autumn 2018
f. Tree algorithms (bagging and boosting both) like random forest are also important. You can check this StatQuest channel for that. StatQuest with Josh Starmer
g. A course which you may refer to : Machine learning for intelligent systems from cornell. Lecture videos as well as notes are available on the website
Course Page | Machine Learning for Intelligent Systems .

4. ***Deep Learning***
   a. Practical Deep Learning for Coders
   b. Deep Learning Book
   c. MIT Deep Learning 6.S191
   d. CS231n Winter 2016 Lectures
   e. Natural Language Processing with Deep Learning (Winter 2017) - Stanford NLP with DL
   f. Start with the 5 courses offered by Deeplearning.AI on Coursera - deeplearning.ai

Questions for practice

- 40 Interview Questions for Statistics and Data Science:
  40 Questions On Statistics
- 40 Questions to test a Data Scientist on Machine Learning:
  40 Questions on Machine Learning
- 45 Questions to Test a Data Scientist on the Basics of Deep Learning:
  45 Questions on Deep Learning
- 12 Frequently Asked Questions on Deep Learning: Deep Learning FAQ | Frequently Asked Question Deep Learning

Projects

- The most important thing is your projects. You must know everything about your project from data collection to results and methodology. It's important how well you can frame your project in a storyline keeping the interviewer excited in every bit and piece.
- Total 2-4 projects are enough
- If you are focused on data science profile go for only ml projects
- If you are going for sde profile a mix of both will be good.
- Go for quality projects and aim to create ~3 bigger projects that will impress the interviewers. Here are some tips that you can follow:

- Find a real-world dataset that requires a lot of preprocessing and EDA
- Make your code modular: create separate classes for models, data preprocessing, and end-to-end pipelines
- Use test-driven development (TDD) while developing a packaged code
- Work with Git and continuous integration services such as CircleCI
- Expose the model's API to the user, e.g. Flask for Python
- Document the code using Sphinx and adhere to code styling guidelines (e.g. PEP-8 for Python)

DSA Knowledge

Even if you want an ML internship it's recommended to do a decent amount of coding.
Leetcode and interviewbit are good places to develop and enhance required skills

- Leetcode:-- LeetCode
- InterviewBit:-- InterviewBit: Coding Interview Questions

As of now, you can learn the concepts and practice on leetcode, sometimes codeforces too. Keep the interview bit site for the end preparation in summer holidays (it will take approx 2.5 - 3 months)
Other resources and topics are :

- Puzzles from gfg
- Segment Trees, Fenwick Trees etc are less asked topics
- Dp, Geedy, Graphs, Trees, Binary Search, Arrays are important topics
- LeetCode questions asked by Companies

***