My curriculum

1. Resume
   1. 1 page resume
   2. Interview story - walk through resume
   3. Background questions:
      1. Explain past work
      2. Which DS techniques I used?
      3. What languages/tools I am familiar with?
      4. What is the business impact of my work?
2. Technical Interview
   1. Programming Concepts
      1. Data Structures
      2. Algorithms
   2. Probability and Statistics
      1. Probability – basic laws, random variables, discrete distributions, continuous distributions, covariance and correlation, independence, joint, marginal, conditional, mixed distributions, bayes law
      2. Statistics – tests (z, t, chi-square), hypothesis testing, p-value, MLE, Confidence intervals
   3. Machine learning
      1. General predictive modelling – dataset split, pre-processing, feature selection, model selection and evaluation, metrics, hyper-parameter optimization, ensembling
      2. Data exploration and wrangling – handling missing data, noisy data, creating new features , feature transformation, imbalancing
      3. Regression – linear regression, ridge and lasso regression
      4. Classification – logistic regression, KNN, decision trees, random forests, GBM, SVM, NN
      5. Clustering - K-means, Mean-shift, Mixture models, GMM, EM
      6. Dimensionality reduction – PCA, factor analysis, mixture models, ICA, manifold
      7. Other topics – regularization, bagging and boosting, best ML practices
   4. Deep learning
      1. Basics – why and where DL needed, shallow v/s deep networks, activation functions, initializations methods, Bias/variance tradeoff, cost functions, loss/accuracy curves, optimizers (GD, adam, rmsprop, adagrad), regularization (dropout, batch normalization, data augmentation, early stopping, L1, L2), vanishing/exploding gradients, gradient checking, weight decay
      2. Training NN – problem framing, data pre-processing, architecture, hyper-parameters optimization, babysitting the learning process, reproducibility and saving the models, model evaluation criteria
      3. Classification models – CNN and various architectures (LeNet, ResNet, AlexNet, VGG, Googlenet, Siamese etc.)
      4. Time-series models – RNN, LSTM, GRU, image captioning
      5. Generative models – PixelRNN/CNN, AE , GAN
      6. Object Detection – basics, YOLO, SSD
      7. Semantic Segmentation – basics, FCN, U-Net, SegNet
      8. Visualization and understanding – GradCAM, features, filter pattern, activation maps, saliency map
      9. Other topics- transfer learning and multi-task learning, handling data imbalancing
   5. Image Processing and computer vision
   6. Tools/frameworks
      1. Python, C/C++
      2. Numpy
      3. Pandas
      4. Sklearn
      5. Keras
      6. Tensorflow
      7. GDAL
      8. Git
      9. SQL? ggplot?
   7. Puzzles
   8. Case studies - define the problem, solve it and explain its business impact

Resources

1. Programming – Notes, codeschool.io,
2. Probability and stats – MIT course, notes from Suraj, past notes of IIT class
3. ML - Andrew NG, Machine learning case study approach, HOML, some part form Univ of michicgan course
4. DL - Andrew NG, CS231, HOML
5. IMP/CV - BKM Sir notes, some notes form Mubarak Shah and previous courses
6. Tools – sendtex for python, CS50 for C, CS231 for numpy and pandas, HOML for sklearn and tensorflow, keras documentation

Other tasks:

1. Create website and update linkedIN
2. Maintain Github – code whatever I learn
3. Prepare interview question bank topic-wise
4. Be active on twitter

Stats and prob- stats modelling

Programming

Ml tecnhiuqs

Business thinking

Culture/role fit

Past work/interview story

7 important branches:

Basic programming language- R/python, pandas, numpy, SQL

Statistics- null hypothetis, p-value, MLE, confidence intervals

Machine Learning- KNN, Random Forsts, ensemble methods

Data wrangling, exploration, transformation and cleaning- clean up data, identifying corrupt data and then correcting it, handling missing and corrupted data

Data visualization- ggplot

Software engineering- DS and algos run-time and use-cases

Product management – selecting right metrics to select your product eg. usability testing, wireframing, retention and conversion rates, traffic analysis, customer feedback, internal logs, A/B testing (optional, can be said as future work)

Interview questions- Background or technical

Background – walk through resume, explain past work, which techniques I used, languages/tools I am familiar with

Technical-

1. Regression- p-value, interpreatation of coefficients, interaction coefficients, residuals, assumptions, logistic regression
2. General predictive modelling- data prep, cross-validation, parameter selction, predictor selection
3. Random forest- Decision trees and all basics, pruning, variable importance, gbm
4. Matrix factorization- pca, factor analysis
5. K-means- loss function, algorithm iteration steps, right number of clusters, is optimization convex
6. Stats- t-test, z-scores, chi-square, covariance and correlation questions, prob distribution, hypotehesis testing
7. Time-series

p-values, regularization( L1, L2 and when is it needed, geometric shape pf functions), gradient descent, eigen value (linear algebra basics), bayes’ law, time series, multicollinearity, box-cox transformation, clustering

sitatuonal questions- technical knowledge + to test the ability to communicate your approach to a non-technical manager

Basics and simple concepts are very very important, be confident, creativity is the most important, it’s ok if you don’t know but you should know how to think logically and clearly to solve the problem, approach is important

Case studies- define the problem, solve it and explain its business impact