<http://www.mitbbs.com/article_t/JobHunting/33120253.html>

本着国人互助以及传递正能量的真理，发一下我个人找工作过程中整理的machine   
learning相关面经以及一些心得总结。楼主的背景是fresh CS PhD in computer   
vision and machine learning, 非牛校。  
  
已经有前辈总结过很多machine learning的面试题(传送门： http://www.mitbbs.com/article/JobHunting/32808273\_0.html)，此帖是对其的补充，有一小部分是重复的。面经分两大块：machine learning questions 和 coding questions.  
  
Machine learning related questions:  
-  Discuss how to predict the price of a hotel given data from previous   
years  
-  SVM formulation  
-  Logistic regression  
-  Regularization  
-  Cost function of neural network  
-  What is the difference between a generative and discriminative algorithm  
-  Relationship between kernel trick and dimension augmentation  
-  What is PCA projection and why it can be solved by SVD    
-  Bag of Words (BoW) feature  
-  Nonlinear dimension reduction (Isomap, LLE)  
-  Supervised methods for dimension reduction  
-  What is naive Bayes  
-  Stochastic gradient / gradient descent  
-  How to predict the age of a person given everyone’s phone call history  
-  Variance and Bias (a very popular question, watch Andrew’s class)  
-  Practices: When to collect more data / use more features / etc. (watch   
Andrew’s class)  
-  How to extract features of shoes  
-  During linear regression, when using each attribute (dimension)   
independently to predict the target value, you get a positive weight for   
each attribute. However, when you combine all attributes to predict, you get  
some large negative weights, why? How to solve it?  
-  Cross Validation  
-  Reservoir sampling  
-  Explain the difference among decision tree, bagging and random forest  
-  What is collaborative filtering   
-  How to compute the average of a data stream (very easy, different from   
moving average)  
-  Given a coin, how to pick 1 person from 3 persons with equal probability.  
  
  
Coding related questions:  
-  Leetcode: Number of Islands  
-  Given the start time and end time of each meeting, compute the smallest   
number of rooms to host these meetings. In other words, try to stuff as many  
meetings in the same room as possible  
-  Given an array of integers, compute the first two maximum products(乘积)   
of any 3 elements (O(nlogn))  
-  LeetCode: Reverse words in a sentence (follow up: do it in-place)   
-  LeetCode: Word Pattern  
-  Evaluate a formula represented as a string, e.g., “3 + (2 \* (4 - 1) )”  
-  Flip a binary tree  
-  What is the underlying data structure for JAVA hashmap? Answer: BST, so   
that the keys are sorted.  
-  Find the lowest common parent in a binary tree  
-  Given a huge file, each line of which is a person’s name. Sort the names  
using a single computer with small memory but large disk space  
-  Design a data structure to quickly compute the row sum and column sum of   
a sparse matrix    
-  Design a wrapper class for a pointer to make sure this pointer will   
always be deleted even if an exception occurs in the middle  
-  My Google onsite questions: http://www.mitbbs.com/article\_t/JobHunting/33106617.html  
  
面试的一点点心得：  
最重要的一点，我觉得是心态。当你找了几个月还没有offer，并且看到别人一直在版  
上报offer的时候，肯定很焦虑甚至绝望。我自己也是，那些报offer的帖子，对我来说  
都是负能量，绝对不去点开看。这时候，告诉自己四个字：继续坚持。我相信机会总会  
眷顾那些努力坚持的人，付出总有回报。  
machine learning的职位还是很多的，数学好的国人们优势明显，大可一试, 看到一些  
帖子说这些职位主要招PhD，这个结论可能有一定正确性。但是凭借我所遇到的大部分  
面试题来看，个人认为MS或者PhD都可以。MS的话最好有一些学校里做project的经验。  
仔细学习Andrew Ng在Coursera上的 machine learning课，里面涵盖很多面试中的概念  
和题目。虽然讲得比较浅显，但对面试帮助很大。可以把video的速度调成1.5倍，节省  
时间。  
如果对一些概念或算法不清楚或者想加深理解，找其他的各种课件和视频学习，例如  
coursera，wiki，牛校的machine learning课件。  
找工作之前做好对自己的定位。要弄清楚自己想做什么，擅长做什么，如何让自己有竞  
争力，然后取长补短（而不是扬长避短）。  
感觉data scientist对coding的要求没有software engineer那么变态。不过即便如此  
，对coding的复习也不应该松懈。  
  
  
我个人觉得面试machine learning相关职位前需要熟悉的四大块：  
Classification:  
Logistic regression  
Neural Net (classification/regression)  
SVM  
Decision tree  
Random forest  
Bayesian network  
Nearest neighbor classification  
  
Regression:  
Neural Net regression  
Linear regression  
Ridge regression (add a regularizer)  
Lasso regression  
Support Vector Regression  
Random forest regression  
Partial Least Squares  
  
Clustering:  
K-means  
EM  
Mean-shift  
Spectral clustering  
Hierarchical clustering  
  
Dimension Reduction:  
PCA  
ICA  
CCA  
LDA  
Isomap  
LLE  
Neural Network hidden layer  
  
最后祝各位好运。那些还在继续找工作的亲们，坚持住，加油！