

# Part 1. 数据结构与算法相关

1. 请实现**快速排序**算法，自行设计测试用例来说明算法的准确性，算法的时间和空间复杂度是多少？最坏情况下时间复杂度多少？
2. 请实现**归并排序**算法，自行设计测试用例来说明算法的准确性，算法的时间和空间复杂度是多少？最坏情况下时间复杂度多少？
3. 面对一个具体的问题，你倾向于使用归并排序还是快速排序，为什么？

4. 对于输入大小  $n$ , 以下每个数代表对应的时间复杂度, 请按照从小到大的顺序排一下
- $2^{\log n}, \sqrt{2}^{\log n}, n^2, n!, (\log n)!, n^3, \log(n!), n * 2^n, n \log n, e^n, \log n, 1, 2^{2^{n+1}}$

5. 利用主定理来回答以下几个算法的时间复杂度:

$$T(n) = 3 * T\left(\frac{n}{3}\right) + n$$
$$T(n) = 4 * T\left(\frac{n}{2}\right) + n^{2.5}$$

6. 计算递归函数的空间复杂度。假如有一个序列是 1, 1, 1, 3, 5, 9, 17.... 利用递归来表示的话:  $T(n) = T(n-3) + T(n-2) + T(n-1)$ , 如果用此递归来实现第  $N$  个数的求解过程, 此算法的空间复杂度是多少?

7. 我们有  $N$  个没有排序的数列  $a_1, a_2, \dots, a_n$ , 现在需要从中选择最大的  $K$  个数 ( $K < N$ ), 请给出实现算法, 确保算法的复杂度为  $O(N \log K)$ . (hint: priority queue)
8. 对于给定的整数  $n > 0$ , 请求出它的平方根 ( $\text{sqrt}(n)$ ), 保留小数点后一位, 不要使用任何 built-in function. (hint: binary search)
9. 目前有两个已经排序好的数组(从小到大)  $a_1, a_2, \dots, a_n$ ,  $b_1, b_2, \dots, b_m$ , 请求出两个数组合并之后的 median。尽可能想出最高效的方法

10. array 和 linked list 在数据结构中的区别是什么？在查找，删除，增加操作中的复杂度分别是多少？

11. 什么 hash table？它的优点是什么？它的缺点是什么？什么叫 collision？怎么解决？

12. 给定已经排序好的数组  $a_1, a_2, \dots, a_n$ ，请返回其中的两个数，保证它们的之和为 K（整数）。如果不存在，则返回-1。例子：1,4,6,8,10, K=10，需要返回去 4,6

13. 我们平时都接触过 short url, 比如对于 <https://greedy.com/154328794837567/685949281810>, 我们可以存储成 <https://greedy.com/afsik>. 这种 URL 有很多好处 (节省传输 band width, 隐私等等)。假设我们要设计一个类似的服务来接受 short url, 然后 redirect to long url, 而且我们需要处理上百亿个这种 url. 如果你是系统架构师会怎么设计? (hint: distributed hash)
14. 树形的数据结构有很多种, 包括 binary search tree, red-black tree, B-tree 等等。是否了解在数据库系统中常用的数据结构是哪一种? 为什么?
15. 给定两个字符串 s1 和 s2, 请计算它们的编辑距离。编辑距离中我们默认有 3 种不同的操作, 分别是 add, delete, 以及 replace, 每个操作的 cost 为 1.

16. 给定一个数组（有正，有负），找出合为最大的连续的子数组（也叫做 maximum subsequence problem）。比如对于一个数组 1, 5, -4, 3, 6, 3, -3, -5, 1, 合为最大的子数组为(1, 5,-4, 3, 6, 3)

17. 给定一个字符串，请写出它所有可能的 permutation。例子，给定 s="abca"，它可能的所有字符串为：“aabc”，“aacb”，“abac”，“abca”，“acba”，“acab”，“baac”，“baca”，“bcaa”，“caab”，“caba”，“cbaa”

18. 什么叫贪心算法，它与动态规划的主要区别是什么？

19. 什么叫 Hamming Distance? 它的主要应用场景是什么?

20. 给定一个有向图  $G$ , 请写程序来遍历图中所有的点, 分别利用 breath first search(BFS)和 depth first search(DFS)。说明如何判断一个图是否包含 Cycle?

21. 请说出什么是 P, NP, NP-hard 以及 NP-complete 问题? 它们之间的关系是怎么样的?



## Part 2. 数学基础

22. 给定两个矩阵，怎么计算它们之间的相乘？怎么计算一个矩阵 inverse？

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}, B = \begin{bmatrix} 4 & 1 \\ 2 & 4 \end{bmatrix}. \quad A*B = ? \quad A^{-1} = ?$$

23. 怎么计算一个向量的 norm？ $a = (3, 1, 5, 1)$ ,  $|a| = ?$

24. 什么是 Frobenius norm？给定  $A = \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$ ，请计算  $\|A\|_F^2$

25. 什么叫矩阵的 determinant? 怎么计算一个矩阵的 determinant?  $A = \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$ .  
 $\det(A) = ?$

26. 什么叫 underflow, 什么叫 overflow? 对于很多的 AI 问题, 如果出现很多概率的相乘, 我们通常都在最前面加  $\log$ , 为什么? 比如  $\operatorname{argmax} p(x)$ , 通常求解  $\operatorname{argmax} \log p(x)$ .

27. 什么叫信息熵? 什么叫互信息? 他们具体的含义是什么?

28.对于 softmax 函数, 我们去实现的时候怎么避免 underflow 或者 overflow?

29.怎么判断一个函数是否是凸函数? 比如  $\log x$ ,  $Ax+b$ ,  $|Ax-b|^2$  是凸函数吗?

30.利用 chain rule 来分解联合概率  $p(x_1, x_2, x_3, x_4, x_5) =$

31. 什么是 KL-Divergence, 写出它的数学细节, 什么时候需要用到它?

32. 什么叫 PSD 矩阵? 怎么判断一个矩阵是 PSD 矩阵? PSD 矩阵跟判断凸函数有什么关系? (hint: Second Order Convexity Condition)

33. 为什么我们需要关注函数是否是凸函数? 主要原因是什么?

34. 线性回归目标函数是否是 Quadratic Programming? 它是否是凸函数?

35. 如果一个函数不是凸函数, 它是否意味着就没有全局最优解? 如果是, 那这时候怎么获得更好的局部最优解? 在 NLP 领域有哪些例子是为了获得更好的局部最优解而设计的?

36. 什么叫 Semi-definite Programming, 请举一个例子来说明, 以及求解的思路。

37.对于有限制条件的优化问题，一般怎么去解决？至少说出 2 个不同的方法。

38.什么叫 Projected Gradient Descent？请利用 Non-negative Matrix Factorization 来说明如何应用它？

39.Linear Programming 的常见的应用场景有哪些？请说出至少 2 种可以用来解决 LP 问题的算法名称。

40.请使用 Linear Programming 来求解 Maximum Flow Problem, 写出详细的过程。

41. 对于 NP Hard 问题, 我们一般可以使用近似求解方法, 其中问题的松弛化是一种常见的方法。请描述如何使用松弛化技术来求解 Set Cover Problem? 它的近似度是多少?

42.Integer Programming 是否是凸函数问题? 一般如何求解?



## Part 3. 机器学习基础

43.老板给了你一个关于癌症检测的数据集，你构建了二分类器然后计算了准确率，为 98%，你是否对这个模型很满意？为什么？如果还不算理想，接下来该怎么做？

44.怎么判断一个训练好的模型是否过拟合？如果判断成了过拟合，那通过什么办法可以解决过拟合问题？

45.对于线性回归，我们可以使用 Closed-Form Solution，因为可以直接把导数设置为 0，并求出参数。在这个 Closed-Form 里涉及到了求逆矩阵的过程，什么时候不能求出其逆矩阵？这时候如何处理？

46.关于正则，我们一般采用 L2 或者 L1, 这两个正则之间有什么区别？ 什么时候需要用 L2， 什么时候需要用 L1？

47.正则项是否是凸函数？ 请给出证明过程。

48.什么叫 ElasticNet？ 它主要用来解决什么问题？ 具体如何去优化？

49. 基于 Coordinate Descent 算法给出 LASSO 的优化推导过程。

50. 请推导逻辑回归模型：目标函数的构建，最优解的求解过程（SGD）需要详细写出。

51. 在数据线性可分的情况下，为什么逻辑回归模型的参数会变得无穷大？怎么避免？

52.逻辑回归是线性还是非线性模型？为什么？请给出推导过程。

53.我们在使用逻辑回归模型的时候，通常把连续性变量切分成离散型变量，为什么？有什么好处？

54.朴素贝叶斯应叫 Naïve Bayes, 请说出朴素贝叶斯模型的构建过程以及预测过程，并说出为什么叫“naive”？

55.什么叫生成模型，什么叫判别模型？朴素贝叶斯，逻辑回归，HMM，语言模型中哪一个是生成模型，哪一个是判别模型？

56.决策树与随机森林的区别是什么？如果让你选择，你会使用哪个模型，为什么？

57.请介绍 k-means 算法的流程，写出 k-means 模型的目标函数。K-means 求解过程跟 EM 算法之间有什么关系？K-MEANS 目标函数是否是 convex？

58. 什么叫 EM 算法？有哪些经典模型的求解过程会用到 EM 算法？

59. EM 算法是否一定会收敛？EM 算法给出的全局最优还是局部最优？

60. 请解释什么叫 MLE，什么叫 MAP？请说明它们之间的区别。在数据量无穷多的时候，是否 MAP 趋近于 MLE 估计？

61. 请解释什么叫召回率，精确率，F1 Measure，ROC，AUC？什么时候需要用到这些？

62. 数据集拥有非常多的特征，但样本个数有限，所以计划做特征选择，有哪些方法可以用来做特征选择呢？

63. 随机森林和 Gradient Boosting Tree 之间的区别是什么？



64.在构建决策树模型的时候，我们一般不会构建到底，也就做一些剪枝的操作，为什么？然而，在构建随机森林的时候剪枝的操作不像决策树里那么重要，为什么？

65.什么样的数据是不均衡数据 (imbalance data) ？ 需要怎么样的处理？

66.什么是 kernel trick？ 它有什么好处？ 并写出 RBF kernel, Gaussian Kernel 的公式。

67. 什么 Mercer's Theorem, 阐述一下具体的细节。

68. 使用非线性 Kernel 有哪些优缺点? 重点介绍一下效率上的缺点, 并说明为什么会  
产生效率上的缺点?

69. SVM 是 margin-based classifier, 试着推导 SVM, 并说明什么是 KKT 条件。

70. 如果不考虑 kernel, 逻辑回归和 SVM 的区别是什么?

71. 在随机梯度下降法里怎么有效地选择学习率? 有哪些常见的动态改变学习率的策略?

72. 深度学习是什么? 它跟所谓的传统的学习模型有什么本质的区别? 从模型的 Capacity, Hierarchical Representation 的角度举例说明。

73.PCA 的原理是什么？ 推导一下 PCA 的过程。

74.什么叫 PAC 理论？ 它主要用来解决什么问题？

75.解释一下矩阵分解算法以及怎么用到推荐系统里，并利用梯度下降法来推导矩阵分解过程。

76.模型参数和超参数的区别是什么？

77.什么叫因变量，以及因变量模型？

78.超参数的选择方法有哪些？至少列出 4 种以上来说明，并说出其优缺点。

79.什么是 XGBoost 模型？说明一下其技术细节。

80.怎么把 K-means 算法应用到大规模的数据上？有什么 Scalable 的方法？(hint: mini-batch, triangle inequality)

81. K-means 算法与 GMM 之间有什么关系？

82.在深度学习模型里，有哪些技术可以用来避免过拟合现象？

83.CNN 里面 POOLING 的作用是什么？ 卷积的作用是什么？

84.在分类问题里，最后一层通常使用 softmax，请写 softmax 函数。

85.描述一下 SGD, Adagrad, Adam 算法之间的区别, 什么时候使用 SGD? Adagrad? Adam 算法?

86.简单描述一下什么是 Variational Autoencoder(VAE), 什么是 Generative Adversarial Network(GAN)

87.Dropout 和 Bagging 模型的关系是什么? 为什么 Dropout 可以起到避免过拟合的作用?



88.对于拥有两层隐含层的神经网络（MLP），请手动推导其 BP 算法的细节。

89.什么叫 unsupervised Layer-wise Pre-training? 以及说出为什么它可以起到更好初始化的效果? 从优化 Variational Lowerbound 的角度来说明。

90.什么叫表示学习里的 Disentanglement? 利用 NLP 的例子来说明它的一个具体场景。

91. 好的表示(good representation)需要具备哪些特点？ 请说出至少 4 个方面的特点。

92. PyTorch 与 Tensorflow 之间主要区别是什么？ 他们各自的优缺点是什么？

93. 使用 KNN 会遇到一些效率上的问题，请说明如何使用 LSH (latent semantic hashing) 来做近似操作？

## Part 4. 自然语言处理

94.中文分词使用过哪些工具？这个工具里的分词功能具体是怎么实现的？ 比如 Jieba 的算法是怎么实现的？

95.给定两个文本，怎么计算他们之间的相似度？ 欧式距离和余弦相似度之间的主要区别是什么？

96.请使用 Python 去实现一个倒排表。

97.在信息检索领域我们会经常做信息去重，包括在写爬虫的时候，很多爬虫系统都有 Deduplication 的功能。请说明如何实现去重的功能？

98.文本中可以提取的特征有哪些？请把所有能想到的特征描述一下。

99.现在我们想要搭建一个针对于每一个商品特征的情感分析系统，请说出一个可行的方案。

100. 在朴素贝叶斯和语言模型中我们通常会使用 smoothing 技术，请简述几个常见的 smoothing 方法以及它们优缺点。

101. 什么叫 N-gram? 随着 N 值逐渐变大的时候，会出现什么问题？这个问题可以怎么解决？

102. 什么叫 Partition Function? 为什么 Partition Function 比较难计算？对于连续性随机变量，partition function 如何处理？

103. 什么叫 Detail Balance? 请说明为什么 Markov Chain 会收敛? 并请举例说明为什么 PageRank 算法会收敛?

104. 什么叫因变量模型中常说的 complete case 和 incomplete case?

105. 请写出 HMM 中用来做推理(inference)的 Viterbi 算法。它给出的解是否是全局最优解? 为什么?

106. 什么叫 log-linear 模型，它与 CRF 之间有什么关系？

107. 简单描述一下 HMM 和 CRF 的区别，以及 HMM 中的参数学习过程和 CRF 中的参数学习过程。

108. 请选择下面其中一个算法，并做详细的描述：Snowball, KnowitAll, RunnerText。



109. 什么叫实体消歧，一般在文本里怎么解决消歧问题？

110. 什么叫实体统一？ 怎么解决实体统一问题？ 请举例说明如何做地址的统一？

111. 怎么去存储一个知识图谱？ RDF 与属性图之间的区别是什么？

112. 如果我们的图谱里有属性，如何使用 RDF 来存储这些信息？

113. 什么叫分布式表示法，它与 One-hot Encoding 之间有什么区别？

114. 描述一下 Skip-gram 模型，并说明它跟 CBOW, Glove 之间的区别，具体有什么优缺点？详细推导一下 Skip-gram 模型。

115. 学习词向量的时候，我们既可以使用矩阵分解的方法也可以使用 skip-gram 等方法，它们之间的区别是什么，优缺点？

116. 如果在实际项目中遇到了 OOV(out of vocabulary)情况，我们需要如何处理？

117. 请描述什么叫 ELMo，它是如何做到动态学出每一个单词在上下文中的表示？

118. 用分布来 (distribution) 表示一个词相比用一个向量来表示有什么优势? 请使用 Gaussian Embedding 的例子来说明。

119. 什么是 Wasserstein Distance? 它与 KL-Divergence 最大的区别在哪里?

120. 什么时候可以使用非欧式空间来做表示学习? 能描述一下什么叫黎曼空间吗?

121. 什么是 T-SNE, 为什么可视化词向量的时候不使用 PCA, 而使用 TSNE?

122. 学出来的词向量怎么去评估? 有哪几种评估方法?

123. 请阐述如何实现聊天机器人中的 NLU 部分? 分别说明实现意图分类以及实体抽取的细节。

124. 请说明 TransE 的详细过程以及推导细节。它如何应用到图挖掘领域的应用？比如 like prediction？

125. 在 Node2Vec 中它是如何结合广度/深度优先算法和 SkipGram 模型？如果你接着改进模型，会如何改进？

126. 什么叫 Adversial Learning？能否举例说明一下它的应用场景？

127. 什么叫 Privacy-Preserving Learning? 能否举例说明一下它主要的应用场景?

128. 任务导向型聊天机器人如何去评估它的性能好坏?

129. 好的表示需要具备一定的稀疏性, 为什么? 能否举例说明?

130. 请用数学的角度描述一下为什么 RNN 会产生梯度消失现象？LSTM 是怎么解决梯度消失问题的？

131. 实现 mini-batch learning 的时候，我们一般都会把数据弄成同样的长度，为什么？

132. 能简述什么叫 Gradient Clipping 吗？它的优缺点是什么？



133. 请描述 LSTM 和 GRU 之间的区别是什么？

134. 什么是 Beam Search？它在 NLP 中的什么场景里会用到？它跟动态规划的区别是什么？

135. 在 Seq2Seq 模型训练好之后做预测时一般会采用 Greedy Decoding 或者 Beam Search，这两者之间有什么区别？

136. 现在想实现一个实体分类的任务，采用的模型是 Bi-LSTM-CRF, 请详细描述模型的细节，以及写出核心的 Pytorch/Tensorflow 代码。

137. 为什么需要注意力机制？它能带来什么样的好处？请说明如何把注意力机制与 Seq2Seq 模型结合使用？

138. 相比基于 LSTM 的 Seq2Seq 模型，Transformer 它的优势在哪儿？它主要用来解决 LSTM 的哪几个缺陷？

139. Self-Attention 中使用的 Q, K, V 矩阵分别代表什么？ 它是如何工作的？ 请画出整个流程。

140. 为什么 Transformer 里使用了 multi-head attention, 多使用几个 Q,K,V 矩阵的优势是什么？

141. 在 Transformer 里我们没有引入任何的时序关系，那单词之间的关系是如何来实现的？ 你觉得这种方式好还是不好？

142. 如何使用 Transformer 搭建一个闲聊型聊天机器人？画出模型架构图。

143. 什么叫 Masked Language Model？它跟一般的 language model 有什么区别？

144. BERT 为什么会使用 Transformer Encoder 来实现预训练，而不是 Decoder 端？

145. 如何使用 BERT 去搭建机器翻译系统，请画出模型架构图，并写出主要代码模块。

146. 如何结合 BERT 和 CRF 实现实体分类的程序？请详细描述一下模型架构，写出核心代码模块。

147. 什么叫 MASS，请解释说明它主要用来解决什么问题。

148. 如何结合 CRF 和 BERT 模型去解决实体分类问题？请描述需要用到的模型架构。

149. 对于 Seq2Seq 模型（机器翻译）如何可视化中间的结果？结合 Layer-wise Relevance Propagation 算法解释说明。

150. 如果把朴素贝叶斯看做是生成模型，它跟 LDA（主题模型）之间主要区别在哪儿？

151. 如果对短文本（比如微博）做聚类，在这个过程当中你预计会遇到什么样的问题？

152. 具体描述什么是 LDA 模型，以及利用 Collapsed Gibbs Sampler 来推导 LDA 模型的参数。

153. 我们都知道 LDA 模型是不考虑单词的先后顺序的，也是经典的 Bag-of-words 模型，也可以看作是 Unigram 模型。如何改造 LDA 使得得到类似于 Bigram 模型？

154. 什么叫 Conjugate Prior? 哪些分布之间拥有这种特点?

155. 什么是变分法? 它与 MCMC 之间有什么主要的区别? 从偏差 (bias), 效率 (efficiency), 分布式 (distributed) 的角度来考虑。

156. 在贝叶斯模型 (Bayesian Model) 里什么叫无参数模型 (nonparametric model)? 对于无参数模型我们需要采取什么样的措施?



157. 什么叫 MCMC 中的 Langevin Dynamics? 它跟 MAP 估计之间有什么区别?

158. 对于吉布斯采样, 怎么判断采样过程已经收敛?

159. 什么叫变分法里的 ELBo, 请写出推导的过程。

160. 请描述一下什么叫 Bayesian Recurrent Neural Network, 详细阐述模型里涉及到的细节。

161. 请描述 Chinese Restaurant Process 的过程, 以及说明如何结合 LDA 模型来处理无参情况?

162. 请使用贝叶斯的思想去解释 Dropout。