## 传统机器学习+深度学习问题

1. 模型评价标准有哪些？\*
2. ROC横坐标和纵坐标分别是什么？\*
3. AUC是什么，画曲线。\*\*\*
4. 样本不均衡是否会对AUC造成影响？\*
5. 知道哪些损失函数？\*
6. 交叉验证的原理和作用？与之间划分训练测试集相比的优点？\*
7. 梯度下降法，牛顿法，拟牛顿法的区别\*\*
8. 为什么梯度是函数变化最快的方向？\*
9. KNN\*
10. LR推导\*\*\*\*\*\*\*\*\*\*
11. LR的缺点\*
12. LR损失函数\*\*\*
13. LR和最大熵模型的相似点和不同点\*
14. 讲一下softmax\*\*
15. 推导softmax的梯度，和tanh的梯度（求导~）\*
16. Softmax的损失函数？\*
17. 解释下交叉熵\*\*
18. 最大似然和交叉熵\*\*
19. 极大似然函数和极大后验函数是啥？\*
20. 回归问题的损失函数都有哪些？从哪些角度设计一个合理的损失函数？\*
21. 哪些场景下的分类问题不适用交叉熵损失函数？\*
22. 手写LR梯度下降过程\*\*
23. LR怎么实现多分类？\*
24. SVM推导\*\*\*\*\*\*\*\*\*
25. SVM的损失函数\*\*\*\*
26. SVM核函数，说几个？对核函数有什么要求？本质是什么？\*\*\*\*\*
27. SVM如何选择核函数？\*\*
28. SVM如何多分类，为什么一般不用？\*\*
29. LR和SVM的区别\*\*\*\*\*\*\*\*\*\*
30. 在数据量比较大的情况下LR和SVM哪个好？\*
31. LR和boost的区别\*
32. LR/Boost和朴素贝叶斯的区别\*
33. svm为什么要转化成对偶形式？\*\*\*\*\*\*
34. KKT条件具体是什么？\*
35. svm什么时候用线性核和高斯核？\*
36. 对VC维的理解\*
37. 生成式模型和判别式模型的区别，分别有哪些？\*\*
38. PCA原理，和LDA的区别\*\*
39. LDA的推导
40. SVD\*
41. 朴素贝叶斯的原理，”朴素”应该怎样理解\*\*
42. HMM和CRF(NLP岗容易问到)\*\*\*\*\*\*(蓝色的三个题一般问到一个另外两个也会问)
43. GMM算法\*\*
44. 讲下EM算法，核心是什么？E步和M步的具体步骤？E中的期望是什么？\*\*\*
45. 相对于传统机器学习，NN有什么优缺点？\*
46. 神经网络为什么用交叉熵损失函数？\*
47. DNN的反向传播机制？推导BP\*\*
48. DNN的梯度更新方式？(优化方法)\*
49. BGD,SGD和Mini-bgd的区别\*
50. 梯度消失和梯度爆炸产生的原因，如何改善\*\*\*\*
51. 相比于DNN，CNN的区别和优点？\*
52. 为什么样本输入前要做归一化/标准化？\*
53. 深度学习如何加快训练过程的收敛速度？\*
54. 深度学习特征的hash搜索算法时间复杂度\*
55. 深度学习中提升模型泛化能力的方法都有哪些？\*
56. 深度学习中L2正则化合dropout有哪些区别？\*
57. LSTM的图和原理推导\*\*\*\*
58. GRU(LSTM变体)介绍\*
59. CNN介绍，卷积的作用，池化的作用\*\*\*\*
60. 1\*1卷积的作用\*
61. 池化层反向传播如何计算？\*
62. Max pooling和average pooling有什么区别？为什么降采样用max分类用average？max pooling是怎么做反向传播的？\*
63. CNN与RNN的区别\*
64. 感知野有什么作用？\*
65. CNN的发展历程，出现过的主流网络结构，解决了什么问题？\*
66. 说一下最近两年新的CNN网络\*
67. 为什么pooling层不好，要用胶囊网络？\*
68. CNN结构未来的发展方向\*
69. 卷积核如何进行权值共享？为什么要权值共享？\*\*
70. 大的卷积核和小的卷积核相比有什么优缺点？为什么现在基本不用大的了？\*
71. ResNet的思想\*\*
72. ResNet旁路跟卷积后的结果相加具体是怎么做的？\*
73. 讲讲ResNet的block结构，为什么要这么做？\*\*
74. Inception的结构，解决了什么问题？\*
75. Inception如何缓解梯度消失？\*
76. 指针网络\*
77. 指针网络/seq2seq的区别\*
78. sequence to sequence\*
79. RNN的结构\*
80. BPTT的推导\*
81. RNN梯度消失问题\*\*
82. RNN梯度爆炸怎么解决？\*
83. LSTM的原理+推导\*\*
84. LSTM里面每个gate分别是干什么的？\*
85. 遗忘门是一个数值还是向量?维度是多少?\*
86. LSTM相比RNN的优点，为什么能解决RNN的梯度消失问题？\*\*
87. LSTM如何处理变长的序列？\*
88. LSTM一个batch里面长度不一怎么办，训练会出现什么问题？\*
89. 对LSTM的attention机制的理解？\*
90. 画出GRU的cell结构，计算公式\*
91. GRU与LSTM的区别，改进了LSTM为什么不用？\*\*
92. LSTM和GRU哪个好？\*
93. CV中data augmentation有哪些？\*
94. Dropout的原理和作用\*\*\*\*
95. Dropout随机选神经元输出0，下一次传播这些神经元权重是随机初始化还是不变？\*
96. Dropout train和dropout test的区别\*
97. Dropout为什么能防止过拟合？\*
98. GAN的原理\*
99. 激活函数的作用\*
100. Sigmoid的缺点是什么/为什么会梯度消失？\*\*
101. relu如何缓解梯度消失？但它并不能彻底解决梯度消失问题，为什么？\*
102. Relu和sigmoid的区别和各自的优点\*\*\*
103. 梯度消失问题和损失函数有关吗？\*
104. 交叉熵损失函数，相比均方误差损失函数的优点\*
105. Batch Normalization的原理和作用\*\*\*\*\*\*\*
106. 为什么Batch Normalization能防止过拟合？\*\*\*
107. BN的训练和测试的区别，讲一下过程。测试的时候均值和方差怎么来的？\*\*
108. 优化算法：Adam的原理\*\*
109. 继上题：二阶矩阵是什么，为什么要用？\*
110. 优化算法：手写SGD伪码\*
111. 如何计算两个文本之间的相似度？\*
112. 介绍word2vec的原理\*\*\*\*\*\*\*
113. Fasttext原理\*
114. 画出fasttext的网络结构，描述其在分类和embedding时的区别？讲讲训练过程。\*
115. Embedding\*
116. 跨语言 word embedding 出现OOV怎么办\*
117. 给定一个query和一个doc，能想到哪些衡量相关性的方法？\*
118. 训练误差与验证误差的关系\*
119. 过拟合和欠拟合\*\*
120. 如何判断过拟合\*\*
121. 防止过拟合的方法\*\*\*\*\*\*\*\*
122. 讲讲机器学习中的方差和偏差，对于训练的意义？\*\*\*
123. 讲讲bagging和boosting各自对于偏差方差的侧重点\*
124. 方差偏差分解的公式\*
125. Kmeans原理和训练过程\*\*\*
126. Kmeans的时间复杂度和空间复杂度\*
127. 距离的计算方法，向量之间？字符串之间？\*\*
128. DBSCAN的原理和算法伪码，与Kmeans, OPTICS的区别\*
129. 决策树原理，剪枝和评价指标\*
130. 决策树种类以及分裂方式(ID3、C4.5、CART)？\*\*
131. CART回归树是怎么实现的\*
132. CART分类树和ID3以及C4.5有什么区别\*
133. 树模型如何处理缺失值和连续值？\*
134. 决策树分类和决策树回归有什么区别？\*
135. 决策树如何选择划分节点？\*\*
136. 决策树做回归的时候划分点怎么做选择？\*
137. 为什么有了信息增益还提出信息增益比？\*
138. 信息增益、信息增益比、基尼系数的公式和原理\*
139. 决策树对数据的划分进行了更细的粒度，这样容易造成数据的过拟合，有什么解决办法？\*
140. 剪枝有哪几种方式？\*
141. 随机森林原理，它和决策树的关系\*\*
142. 随机森林是如何选择特征的？如何评价特征重要性？\*\*\*
143. 随机森林如何防止过拟合，随机体现在哪？\*
144. 随机森林和GBDT的区别\*
145. 随机森林和GBDT的基分类器区别，如何剪枝\*
146. 随机森林，adaboost和GBDT的联系和区别\*
147. Adaboost的学习策略\*
148. Adaboost和GBDT的区别\*
149. 随机森林的树会不会限制它的生长?GBDT呢？为什么？\*
150. GBDT的原理以及常用的参数\*\*\*\*\*\*
151. GBRT的原理，推导以及适用场景\*\*\*\*\*\*\*
152. GBDT与随机森林的区别在哪，分别用在什么地方？\*\*
153. GBDT为什么用梯度来优化，有什么好处？\*
154. 正负样本比例对LR和GBDT的影响？\*
155. XGBoost和LR相比的优缺点？\*
156. XGBoost跟GBDT比有哪些区别、优点？\*\*\*\*\*\*\*\*
157. XGBoost的优化方法和GBDT的优化方法有什么区别？\*
158. XGBoost跟随机森林的区别和各自的优点？\*\*\*
159. XGBoost详细原理推导\*\*\*\*\*\*
160. XGBoost如何防止过拟合\*
161. XGBoost为什么用二阶泰勒展开？\*\*\*
162. XGBoost中L1正则化是怎么用的？
163. XGBoost的行抽样,列抽样可以起到哪些作用？样本少了不是容易过拟合么，为什么还能防止过拟合？\*
164. XGBoost, LR和SVM中哪个处理数据不平衡问题的能力最强？\*
165. Lightgbm的原理和特性\*\*
166. XGBoost, Lightgbm, GBDT的区别和适用场景？\*\*\*\*
167. XGB和LGB的并行如何实现\*\*
168. 正则化都有哪些方法？\*
169. 正则化和偏差误差的关系？\*
170. 介绍L1,L2正则化及其区别，应用场景\*\*\*\*\*\*\*\*
171. 从数学的角度上解释为什么L2正则化能提高模型的泛化能力\*
172. L1正则化如何求梯度？\*
173. 用梯度解释为什么L1正则化有稀疏作用？\*
174. 集成学习介绍 (boosting,bagging,stacking)，bagging和boosting的差异\*\*\*\*\*\*
175. stacking和blending的区别\*
176. 模型融合有哪些方式，区别及应用场景？\*
177. tfidf的原理是什么，请写出计算公式\*\*

## 预处理与特征工程

1. 对缺失数据应该如何处理？异常值如何处理？数据平滑。\*\*
2. 对长拖尾分布的数据进行log处理的好处。
3. 对于一个取值较多的类别变量在不能onehot的情况下如何使用？\*
4. One-hot编码的优势\*
5. 讲讲特征选择的方法有哪些，应用场景\*\*
6. 样本不均衡的情况下如何处理数据？\*
7. 如何解决稀疏问题？\*
8. 统计特征，交叉特征，转化特征。

## 做推荐方面的专业问题

1. 了解推荐算法么？举例说明一般个性化推荐怎么做的？\*
2. CTR\*
3. LR, FM与FFM\*\*
4. 手写推导FM，并说出FM和FFM的时间复杂度\*
5. GBDT+LR\*
6. FFM对比FM的优化\*
7. DeepFM\*
8. 协同过滤\*\*
9. 冷启动问题\*
10. 协同过滤的itemCF, userCF的区别和适用场景\*
11. 推荐系统中你认为最重要的环节是什么？\*
12. 多臂老虎机中，有许多方法，比如e-greedy，timponson采样，UCB，这些方法都有哪些适用场景？\*

## 语音方面的专业问题

1. MFCC特征维数是多少，有没有进行差分？\*
2. MFCC特征提取过程\*
3. DTW语音识别的过程\*

## NLP方面的专业问题

1. 文本分类常用的算法有哪些？\*
2. tfidf的原理是什么，请写出计算公式\*

## CV岗专业问题

1. Faster rcnn的RPN和ROI pooling\*
2. ssd和faster rcnn，讲讲它们的区别？\*
3. ssd在哪方面比yolo v1检测小物体更有效？\*
4. **继上题**：损失函数为什么用交叉熵而不用平方损失？平方损失用的场景？\*
5. BN层在微调时需要加载预权重吗？\*
6. Mask-rcnn的原理\*
7. OCR\*
8. 如何用神经网络实现OCR？\*
9. Rcnn到frcnn的进化路程，有什么区别和改进？\*

## 分布式问题

1. mapreduce的思想
2. 简述Hadoop和spark的区别

## 关于比赛

1. 比赛中担任的角色
2. 比赛的难点
3. 特征工程是怎么做的？
4. 上分手段最有效的是哪些？