## 传统机器学习

1. 模型评价标准有哪些？\*
2. P和R怎么计算？F1怎么计算？\*
3. ROC横坐标和纵坐标分别是什么？\*
4. AUC是什么，画曲线。\*\*\*\*
5. 样本不均衡是否会对AUC造成影响？\*
6. 知道哪些损失函数？\*
7. 分类为什么用logloss？\*
8. Logloss和AUC的区别？\*
9. 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？\*
10. 梯度下降法，牛顿法，拟牛顿法的区别\*\*\*
11. 为什么梯度是函数变化最快的方向？\*
12. KNN\*
13. KNN复杂度高怎么解决？\*
14. 机器学习里面常见的激活函数有哪些？为什么通常需要0均值？\*
15. LR和线性回归的区别和联系\*
16. LR推导\*\*\*\*\*\*\*\*\*\*\*\*
17. LR如何解决共线性问题，为什么深度学习不强调？\*
18. LR为什么使用sigmoid函数？\*
19. LR的缺点\*
20. LR损失函数\*\*\*
21. LR和最大熵模型的相似点和不同点\*
22. 如果所有样本都是正样本，那么写出来的LR模型的超平面是什么样的？\*
23. 讲一下softmax，要推导\*\*\*
24. 推导softmax的梯度，和tanh的梯度（求导~）\*
25. Softmax的损失函数？\*
26. 解释下交叉熵\*\*
27. 交叉熵损失函数的形式\*
28. 为什么用交叉熵损失函数而不用均方误差函数？\*\*
29. 哪些场景下的分类问题不适用交叉熵损失函数？\*
30. 最大似然和交叉熵\*\*
31. 极大似然函数和极大后验函数是啥？\*
32. 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？\*
33. 哪些场景下的分类问题不适用交叉熵损失函数？\*
34. 手写LR梯度下降过程\*\*
35. LR怎么实现多分类？\*
36. SVM推导\*\*\*\*\*\*\*\*\*\*\*\*
37. SVM的损失函数\*\*\*\*
38. SVM核函数，说几个？对核函数有什么要求？本质是什么？\*\*\*\*\*
39. SVM如何选择核函数？\*\*
40. SVM如何多分类，为什么一般不用？\*\*
41. SVM能否通过梯度下降来求解？\*
42. LR和SVM的区别\*\*\*\*\*\*\*\*\*\*\*\*\*
43. 在数据量比较大的情况下LR和SVM哪个好？\*
44. LR和boost的区别\*
45. LR/Boost和朴素贝叶斯的区别\*
46. svm为什么要转化成对偶形式？\*\*\*\*\*\*
47. KKT条件具体是什么？\*
48. svm什么时候用线性核和高斯核？\*
49. 对VC维的理解\*
50. 生成式模型和判别式模型的区别，分别有哪些？\*\*
51. 降维有哪些方法？\*
52. PCA原理，和LDA的区别\*\*\*
53. LDA的推导
54. SVD\*\*
55. SVD和特征值的关系\*
56. 朴素贝叶斯的原理，”朴素”应该怎样理解\*\*\*\*
57. HMM最成功的应用是什么？\*
58. 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)
59. GMM算法\*\*\*
60. 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？\*\*\*
61. 训练误差与验证误差的关系\*
62. 过拟合和欠拟合\*\*\*
63. 如何判断过拟合\*\*
64. 防止过拟合的方法\*\*\*\*\*\*\*\*\*\*\*\*
65. 讲讲机器学习中的方差和偏差，对于训练的意义？\*\*\*\*
66. 讲讲bagging和boosting各自对于偏差方差的侧重点\*
67. 方差偏差分解的公式\*
68. 正则化都有哪些方法？\*\*\*
69. 正则化和偏差误差的关系？\*
70. 介绍L1,L2正则化及其区别，应用场景\*\*\*\*\*\*\*\*\*\*\*
71. L1正则化有什么好处？\*
72. 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\*
73. L1正则化如何求梯度？\*
74. 用梯度解释为什么L1正则化有稀疏作用？\*
75. Kmeans原理和训练过程\*\*\*\*
76. Kmeans如何选择k的个数？\*
77. Kmeans的时间复杂度和空间复杂度\*
78. 距离的计算方法，向量之间？字符串之间？\*\*
79. DBSCAN的原理和算法伪码，与Kmeans, OPTICS的区别\*\*
80. 决策树原理，剪枝和评价指标\*
81. 决策树种类以及分裂方式(ID3、C4.5、CART)？\*\*
82. CART回归树是怎么实现的\*\*
83. CART分类树和ID3以及C4.5有什么区别\*\*
84. 树模型如何处理缺失值和连续值？\*
85. 如果有一万个地理坐标，转换成1-10000的数，可以用决策树么？\*
86. 决策树分类和决策树回归有什么区别？\*
87. 决策树如何选择划分节点？\*\*
88. 决策树做回归的时候划分点怎么做选择？\*
89. 为什么有了信息增益还提出信息增益比？\*
90. 信息增益、信息增益比、基尼系数的公式和原理\*\*
91. 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？\*
92. 剪枝有哪几种方式？\*
93. 随机森林原理，它和决策树的关系\*\*\*\*\*\*\*
94. 随机森林特征分裂的依据\*
95. 随机森林的参数\*
96. 随机森林是如何选择特征的？如何评价特征重要性？\*\*\*
97. 随机森林如何防止过拟合，随机体现在哪？\*\*
98. 随机森林和GBDT的区别\*
99. 随机森林和GBDT的基分类器区别，如何剪枝\*
100. 随机森林，adaboost和GBDT的联系和区别\*
101. Adaboost的学习策略\*
102. Adaboost是如何更新样本权重的？\*\*
103. Adaboost和GBDT的区别\*\*
104. 随机森林的树会不会限制它的生长?GBDT呢？为什么？\*
105. GBDT的原理以及常用的参数\*\*\*\*\*\*
106. GBDT分类模型拟合的是什么？\*
107. GBDT如何做分类和回归\*
108. GBRT的原理，推导以及适用场景\*\*\*\*\*\*\*\*
109. GBDT与随机森林的区别在哪，分别用在什么地方？\*\*\*
110. GBDT为什么用梯度来优化，有什么好处？\*
111. 正负样本比例对LR和GBDT的影响？\*
112. XGBoost和LR相比的优缺点？\*
113. XGBoost跟GBDT比有哪些区别、优点？\*\*\*\*\*\*\*\*\*\*\*
114. XGBoost的优化方法和GBDT的优化方法有什么区别？\*
115. XGBoost跟随机森林的区别和各自的优点？\*\*\*
116. XGBoost详细原理推导\*\*\*\*\*\*\*\*
117. XGBoost的VC维实多少？决策树呢？\*\*
118. XGBoost如何防止过拟合\*
119. XGBoost为什么用二阶泰勒展开？\*\*\*\*
120. XGBoost中L1正则化是怎么用的？
121. 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？\*\*
122. XGBoost, LR和SVM中哪个处理数据不平衡问题的能力最强？\*
123. Lightgbm的原理和特性\*\*\*
124. XGBoost, Lightgbm, GBDT的区别和适用场景？\*\*\*\*\*\*\*\*
125. XGB和LGB的并行如何实现\*\*
126. 集成学习介绍 (boosting,bagging,stacking)，bagging和boosting的差异\*\*\*\*\*\*\*\*\*\*
127. Bagging为什么能减小方差？\*
128. stacking和blending的区别\*
129. 模型融合有哪些方式，区别及应用场景？\*
130. tfidf的原理是什么，请写出计算公式\*\*\*\*
131. 强化学习和监督学习的区别？\*
132. DQN的基本原理\*
133. DQN模型为什么要做经验回收\*
134. Dueling DQN和DQN有什么区别？\*

## 深度学习

1. 相对于传统机器学习，NN有什么优缺点？\*
2. 神经网络为什么用交叉熵损失函数？\*
3. 神经网络里面损失函数有哪些？\*
4. 神经网络不收敛有什么可能的原因？怎么办？\*
5. Wide&&Deep的原理\*
6. DNN的反向传播机制？推导BP\*\*\*
7. DNN的梯度更新方式？(优化方法)\*
8. 为什么沿着负梯度的方向进行参数更新？\*\*
9. BGD,SGD和mini-bgd的区别\*
10. 梯度消失和梯度爆炸产生的原因，如何改善\*\*\*\*\*
11. 相比于DNN，CNN的区别和优点？\*
12. 为什么样本输入前要做归一化/标准化？\*
13. 深度学习如何加快训练过程的收敛速度？\*
14. 深度学习特征的hash搜索算法时间复杂度\*
15. 深度学习中提升模型泛化能力的方法都有哪些？\*
16. 深度学习中L2正则化合dropout有哪些区别？\*\*
17. CNN介绍，卷积的作用，池化的作用\*\*\*\*
18. 卷积的复杂度\*
19. 1\*1卷积的作用\*
20. 池化的作用\*
21. 池化层反向传播如何计算？\*
22. 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？\*
23. CNN与RNN的区别\*
24. 感知野有什么作用？\*
25. CNN的发展历程，出现过的主流网络结构，解决了什么问题？\*
26. 说一下最近两年新的CNN网络\*
27. pooling层有什么缺点，要用胶囊网络？\*
28. CNN结构未来的发展方向\*
29. 卷积核如何进行权值共享？为什么要权值共享？\*\*
30. 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？\*
31. ResNet的思想\*\*
32. ResNet旁路跟卷积后的结果相加具体是怎么做的？\*
33. 讲讲ResNet的block结构，为什么要这么做？\*\*
34. Inception的结构，解决了什么问题？\*
35. Inception如何缓解梯度消失？\*
36. 指针网络\*
37. 指针网络/seq2seq的区别\*
38. sequence to sequence\*
39. RNN的结构\*
40. RNN中的正则化方法：AR和TAR\*
41. BPTT的推导\*\*
42. RNN梯度消失问题\*\*\*\*
43. RNN梯度爆炸怎么解决？\*\*
44. LSTM的原理+推导\*\*\*\*\*\*\*\*\*\*
45. LSTM里面每个gate分别是干什么的？\*
46. 遗忘门是一个数值还是向量?维度是多少?\*
47. LSTM相比RNN的优点，为什么能解决RNN的梯度消失问题？\*\*\*
48. LSTM如何处理变长的序列？\*
49. LSTM一个batch里面长度不一怎么办，训练会出现什么问题？\*
50. 对LSTM的attention机制的理解？\*
51. 画出GRU的cell结构，计算公式\*
52. GRU(LSTM变体)介绍\*
53. GRU与LSTM的区别，改进了LSTM为什么不用？\*\*
54. LSTM和GRU哪个好？\*
55. CV中data augmentation有哪些？\*
56. Dropout的原理和作用\*\*\*\*\*\*
57. 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？\*
58. Dropout train和dropout test的区别\*
59. Dropout为什么能防止过拟合？\*
60. GAN的原理\*
61. 激活函数的作用\*
62. Sigmoid的缺点是什么/为什么会梯度消失？\*\*
63. 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？\*
64. Relu和sigmoid的区别和各自的优点\*\*\*
65. 梯度消失问题和损失函数有关吗？\*
66. Batch Normalization的原理和作用\*\*\*\*\*\*\*\*
67. 为什么Batch Normalization能防止过拟合？\*\*\*
68. 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？\*\*
69. Early stop对参数有什么影响？\*
70. 优化算法：Adam的原理\*\*
71. 继上题：二阶矩阵是什么，为什么要用？\*
72. 优化算法：手写SGD伪码\*
73. SGD和Adam的区别和联系\*
74. 如何计算两个文本之间的相似度？\*
75. 介绍word2vec的原理，CBOW和Skip-gram的训练过程，负采样\*\*\*\*\*\*\*\*\*\*
76. Fasttext原理，相比word2vec有哪些改进？\*\*\*
77. Glove原理\*
78. 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。\*
79. Sentence-Embedding, Doc embedding\*\*
80. 跨语言 word embedding 出现OOV怎么办\*

## 预处理与特征工程

1. 数据之间如果不是独立同分布会怎样？\*
2. 对缺失数据应该如何处理？异常值如何处理？数据平滑。\*\*
3. 对长拖尾分布的数据进行log处理的好处。
4. 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？\*
5. One-hot编码的优势\*
6. 讲讲特征选择的方法有哪些，应用场景\*\*\*
7. 样本不均衡(类别不平衡)的情况下怎么解决？\*\*\*\*
8. 如何解决稀疏问题？\*
9. 统计特征，交叉特征，转化特征。\*
10. 数据可视化怎么做\*
11. 如何判断特征的有效性？\*

## 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. 个性化推荐的常用模型\*
3. A/B test\*
4. CTR\*
5. LR, FM与FFM\*\*\*
6. FFM在FM基础上有什么改进？如何确定每一个特征所属的field？\*
7. LR和FM的区别？\*
8. FM需要进行交叉特征的选择么？\*
9. 如果在LR中选了一部分特征做交叉后取得了比FM更好的效果，这是为什么？\*
10. 继上题：FM变成DeepFM后效果又超过了LR，为什么？\*
11. 手写推导FM，并说出FM和FFM的时间复杂度\*
12. GBDT+LR\*\*
13. FFM对比FM的优化\*
14. DeepFM\*\*
15. 画出DeepFM或FNN或PNN的图\*
16. 协同过滤\*\*
17. 冷启动问题\*
18. 协同过滤的itemCF, userCF的区别和适用场景\*
19. 推荐系统中你认为最重要的环节是什么？\*\*
20. 模型在线下评估和线上使用时，往往出现线上实际效果不如线下效果的情况，请分析可能的原因。\*
21. 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\*
3. 如何找相似文档/相似图片？\*
4. 如何衡量两个句子的相似度？\*
5. 文本语义相似度衡量soft match\*
6. 给定一个query和一个doc，能想到哪些衡量相关性的方法？\*

## 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？\*
10. Mobile Net V1和V2的区别？\*
11. Inception V3的结构和V2有什么区别？\*
12. 如果要构造一个卷积神经网络，你会考虑哪些方面，从原理方面去讲\*
13. 胶囊神经网络的原理\*

## 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？