## 传统机器学习+深度学习问题

1. AUC是什么，画曲线。\*
2. 梯度下降法，牛顿法，拟牛顿法的区别\*
3. KNN\*
4. LR推导\*\*\*\*\*\*
5. LR损失函数\*
6. LR怎么实现多分类？\*
7. SVM推导\*\*\*\*
8. SVM的损失函数\*\*
9. SVM核函数？对核函数有什么要求？本质是什么？\*
10. SVM如何多分类，为什么一般不用？\*\*
11. LR和SVM的区别\*\*\*\*\*\*\*\*
12. 在数据量比较大的情况下LR和SVM哪个好？\*
13. LR和boost的区别\*
14. LR/Boost和朴素贝叶斯的区别\*
15. svm为什么要转化成对偶形式？\*\*\*
16. svm什么时候用线性核和高斯核？\*
17. 生成式模型和判别式模型的区别，分别有哪些？\*
18. PCA原理，和LDA的区别\*
19. 朴素贝叶斯的原理，”朴素”应该怎样理解\*\*
20. 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)
21. GMM算法\*\*
22. 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？\*\*\*
23. 相对于传统机器学习，NN有什么优缺点？\*
24. DNN的反向传播机制？\*
25. DNN的梯度更新方式？(优化方法)\*
26. 梯度消失和梯度爆炸，如何改善\*\*
27. 相比于DNN，CNN的区别和优点？\*
28. LSTM的图和原理\*\*\*
29. GRU(LSTM变体)介绍\*
30. CNN介绍，卷积的作用，池化的作用\*\*\*
31. CNN与RNN的区别\*
32. 感知野有什么作用？\*
33. CNN的发展历程，出现过的主流网络结构，解决了什么问题？\*
34. 说一下最近两年新的CNN网络\*
35. 为什么pooling层不好，要用胶囊网络？\*
36. CNN结构未来的发展方向\*
37. 卷积核如何进行权值共享？\*
38. 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？\*
39. ResNet的思想\*\*
40. 讲讲ResNet的block结构，为什么要这么做？\*\*
41. Inception的结构，解决了什么问题？\*
42. Inception如何缓解梯度消失？\*
43. sequence to sequence\*
44. RNN的结构\*
45. RNN梯度消失问题\*
46. LSTM里面每个gate分别是干什么的？\*
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. GRU与LSTM的区别，改进了LSTM为什么不用？\*\*
52. LSTM和GRU哪个好？\*
53. CV中data augmentation有哪些？\*
54. Dropout的原理和作用\*\*\*\*
55. Dropout为什么能防止过拟合？\*
56. GAN的原理\*
57. 激活函数的作用\*
58. Sigmoid的缺点是什么/为什么会梯度消失？\*\*
59. 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？\*
60. Relu和sigmoid的区别和各自的优点\*\*
61. 梯度消失问题和损失函数有关吗？\*
62. 交叉熵损失函数，相比均方误差损失函数的优点\*
63. Batch Normalization的原理和作用\*\*\*\*
64. 优化算法：Adam的原理\*
65. 优化算法：手写SGD伪码\*
66. 如何计算两个文本之间的相似度？\*
67. 介绍word2vec的原理\*\*\*
68. 跨语言 word embedding 出现OOV怎么办\*
69. 给定一个query和一个doc，能想到哪些衡量相关性的方法？\*
70. 训练误差与验证误差的关系\*
71. 如何判断过拟合\*
72. 防止过拟合的方法\*\*\*\*\*
73. 讲讲机器学习中的方差和偏差\*
74. 讲讲bagging和boosting各自对于偏差方差的侧重点\*
75. 方差偏差分解的公式\*
76. Kmeans原理和训练过程\*\*
77. 距离的计算方法，向量之间？字符串之间？\*\*
78. DBSCAN的原理和算法伪码，与Kmeans, OPTICS的区别\*
79. 决策树原理，剪枝和评价指标\*
80. 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？\*
81. 随机森林原理，它和决策树的关系\*\*
82. 随机森林是如何选择特征的？\*
83. 随机森林如何防止过拟合，随机体现在哪？\*
84. 随机森林和GBDT的基分类器区别，如何剪枝\*
85. GBDT的原理以及常用的参数\*\*\*\*\*
86. GBRT的原理，推导以及适用场景\*\*\*\*\*
87. GBDT与随机森林的区别在哪，分别用在什么地方？\*
88. GBDT为什么用梯度来优化，有什么好处？\*
89. XGBoost跟GBDT比有哪些区别、优点？\*\*\*
90. XGBoost跟随机森林的区别和各自的优点？\*
91. XGBoost详细原理推导\*\*\*\*
92. XGBoost如何防止过拟合\*
93. XGBoost为什么用二阶泰勒展开？\*
94. XGBoost中L1正则化是怎么用的？
95. 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？\*
96. XGBoost, Lightgbm, GBDT的区别和适用场景？\*\*\*
97. XGB和LGB的并行如何实现\*
98. 介绍L1,L2正则化及其区别\*\*\*
99. L1正则化如何求梯度？\*
100. 用梯度解释为什么L1正则化有稀疏作用？\*
101. 集成学习介绍 (boosting,bagging,stacking)，bagging和boosting的差异\*\*\*\*
102. stacking和blending的区别\*
103. 特征选择的方法\*
104. 模型融合有哪些方式，区别及应用场景？\*

## 特征工程

1. 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？

## 做推荐方面的专业问题

1. CTR\*
2. LR, FM与FFM\*
3. FFM对比FM的优化\*
4. 协同过滤的itemCF, userCF的区别和适用场景

## 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？\*

## 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？