算法面试问题总结：

Q：梯度下降和随机梯度下降

A：应用场景：在数据量很大时，使用随机梯度下降能获得更快的训练速度。

随机（批量）梯度下降SGD原理：每次计算代价函数倒数时只选择一个（或者少数）样本值。这样可以加快训练速度，并且有一定几率避免局部最小值。（避免局部最优的还有模拟退火算法）

Q：朴素贝叶斯（原理）

A：

Q：PCA 奇异值分解SVD

A：SVD（参照Gilebert Strang线性代数第30讲）：矩阵A行空间基向量矩阵V，列空间基向量矩阵U，,（是标准化系数），此式代表的意义是矩阵空间的转换。将V移到等号右边得到，这就是SVD的形式。通过计算矩阵和的特征值和特征向量得到U和V的值。

Q：ICA原理

A：盲源分离Blind Source Separator、向量内积、相对熵

ICA中假设：独立源产生的信号之间在统计学上是独立的。ICA在实践过程中得益于这个假设的反命题：可以提取出来的统计学独立信号产生自不同的物理源头。

PCA和FA提取出的是相互无关uncorrelated的信号，而ICA提取出相互独立statistical independent的信号，独立是比无关更高级的特性，独立必定无关，无关未必独立。

ICA策略

* 独立：由于源信号是独立的而混合信号不是，则从混合信号中提取出的独立信号应当是源信号。
* 正态：由于混合信号是高斯分布而源信号是非高斯分布，则从混合信号中提取非高斯信号应当是源信号。
* 复杂性：由于源信号相对与混合信号有更简单的结构，则从混合信号中提取出简单结构的信号应当是源信号。

盲源分离的解决方法就是从混合信号中提取出尽量符合这三个特性的信号，这些信号就是源信号。

ICA要求混合信号个数必须等于源信号个数，如果混合信号多可以用PCA降

维或直接丢弃一部分达到平衡，如果混合信号小于源信号则无法提取。（如何解释人耳可以从一个混合声音中区分出不同的人）

参考Independent Components Analysis

Q：算法评价指标

A：一般性指标：泛化能力

分类问题：

准确率，分类正确的比例；

精确率，所有预测为P的样本中TP的比例；

召回率，实际正类中TP的比例；

F1值，准确率和召回率的调和平均；

ROC曲线和AUC面积；

见 <http://blog.csdn.net/abcd1f2/article/details/51887148>

Q：K-Means K指的选取

A：业务选择，轮廓系数，SSE拐点

Q：K-Means缺陷

A：如下：

1）只能发现球状簇类

2）k值得选择

3）初始聚类中心的选择

4）对离群点和孤立点敏感。

Q：Maximum Likelihood Estimate和Maximum Posterior Estimate的关系？

A：似然是指样本出现相对于模型参数的可能性，最大似然就是选出最有可能产生样本的参数。但是在MLE中假设了参数是一个固定但未知的值，在MAP中参数是有分布的，它的优化目标函数是log似然和log参数先验分布之和。Log先验分布起到了正则化得作用，在机器学习模型训练过程中能避免估计到的模型发生过拟合。MLE试用与大样本的情况，而MAP适用于小样本的情况。

参考：《统计学习方法》Pg9 《Spoken Language Process》Pg112

Q：聚类算法

A：k-means、DBSCAN、SOM神经网络（原理）

Q：SVM理解，SVM核函数

A：SVM推导、SVM原理、代价参数C的选择（太大导致过拟合《ESL Pg.424》）

SVM推导如下：

SVM旨在找到特征空间中最优的分类平面使得分类平面到两个类最近的点的聚类最大，此平面唯一确定且能产生最优的分类效果。

条件保证了所有的点到分类超平面的距离至少为，目的是找到最大的和对应的参数。为了去掉，将限制条件变为

或

可以看出的值对满足条件的和并无影响，所以假定，得到原来的目的等价于

该约束条件相当于在分类面周围定义了一个大小为的间隔，所以选择能使得间隔最大的和。这是一个带有线性不等式约束的二次函数优化问题。

用来优化和的拉格朗日函数为

使倒数为0，得到

将上述公式带回得到Wolfe对偶函数

该式求解最终得到的解必须满足Karush-Kuhn-Tucker条件，包含公式1,2,3以及4

从KKT条件可以看出

* 如果，则，换句话说，在分类边界上；
* 如果，则，不在边界上。

进一步理解拉格朗日方法、KKT条件、对偶问题或是优化问题，可以参考博客 <http://anie.me/Lagrangian-And-Dual-Problem/>

SVM核函数

多项式核：

高斯核(Radial basis)：

例：二阶多项式核，X有两个属性。（表示内积）

Q：数据分析一般过程

Q：特征工程

Q：L1、L2正则化

A：例：线性回归lasso采用了L1正则化，Ridge Regression采用了L2正则化。两者都是为了缩减系数，以达到减少过拟合的目的。但是L1在参数空间的坐标轴上是不可导的突出点，L2是一个光滑的圆面，这决定了相较于L2，L1正则化过程中有更大可能将系数缩减到0。 与L1一样有筛减参数个数作用的还有L0正则化，这时正则化的惩罚项就是参数的个数。

对正则化的进一步理解中可联系机器学习策略中的贝叶斯估计。