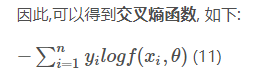
1. 集束搜索（beam search）

主要是用于测试时候进行句子翻译的方法

<https://blog.csdn.net/qq_16234613/article/details/83012046>

1. 交叉熵

主要是用于计算训练模型f(x,θ)中θ的最优值，利用最大似然函数还计算参数组θ的最优值



交叉熵函数可用于多分类multi-classification任务中的损失函数.

CNN 中, 输出层可以采用 softmax 来输出每一类的概率值, 再将 softmax 输出结果, 送入交叉熵损失函数.

Categorical CrossEntropy 多分类的对数损失函数

<https://blog.csdn.net/zziahgf/article/details/79928978>

1. 学习速率

机器学习中的学习速率是指每一步调整的步长，学习速率太小，收敛速度过慢，学习速率太大，会导致代价函数震荡，无法收敛

1. One-hot编码

在机器学习算法中，我们经常会遇到分类特征，例如：人的性别有男女，祖国有中国，美国，法国等。

这些特征值并不是连续的，而是离散的，无序的。通常我们需要对其进行特征数字化。

那什么是特征数字化呢？例子如下：

性别特征：["男"，"女"]

祖国特征：["中国"，"美国，"法国"]

运动特征：["足球"，"篮球"，"羽毛球"，"乒乓球"]

假如某个样本（某个人），他的特征是这样的["男","中国","乒乓球"]，我们可以用 [0,0,4] 来表示，但是这样的特征处理并不能直接放入机器学习算法中。因为类别之间是无序的（运动数据就是任意排序的）。

独热编码

就是讲某个分类变成0,1的二进制向量，比如【男，女】，男的话用【1,0】表示

使用独热编码计算距离会合理些，主要是针对于不好寻找距离意义的情况

<https://www.imooc.com/article/35900>

1. DBSCAN

DBSCAN 算法是一种基于密度的聚类算法：

　　1.聚类的时候不需要预先指定簇的个数

2.最终的簇的个数不确定

<https://www.cnblogs.com/bonelee/p/8692336.html>

1. 神经网络输入要归一化
2. 假设一个神经元有两个输入分别是x1和x2，权重分别是w1和w2，那么该神经元的信号加权求和为x1w1+x2w2。再假设x1属于[0~1]，x2属于[100~1000]，那么x2远远大于x1，那么x1w1就可以忽略不计，整个加权求和就只由x2w2来决定，小的信号就被淹没了！
3. 由于激励函数取值一般都是[0~1]或者是[-1~1]，函数曲线两头趋于直线，无论输入信号数据多么的大，最后取值基本不变，所以输入值太大并没有意义，而且还会让训练速度变得更慢