# CrossEntropy

香农认为：信息是用来消除不确定性的东西，不确定性的数学表达可以转化为概率的形式

信息越少，可能性越大

## 信息熵

把信息中排出了冗余后的平均信息量称为“信息熵”，设表示样本空间的第i个样本，其概率为，样本空间总共有n个样本，信息熵的公式

相关性质：

极值性：在各个样本等概率出现的条件下取最大值，也就是说明等概率的时候最为混乱，而且样本数据越多，熵越大

连续性

对称性（交换样本顺序，结果不变）

## 交叉熵

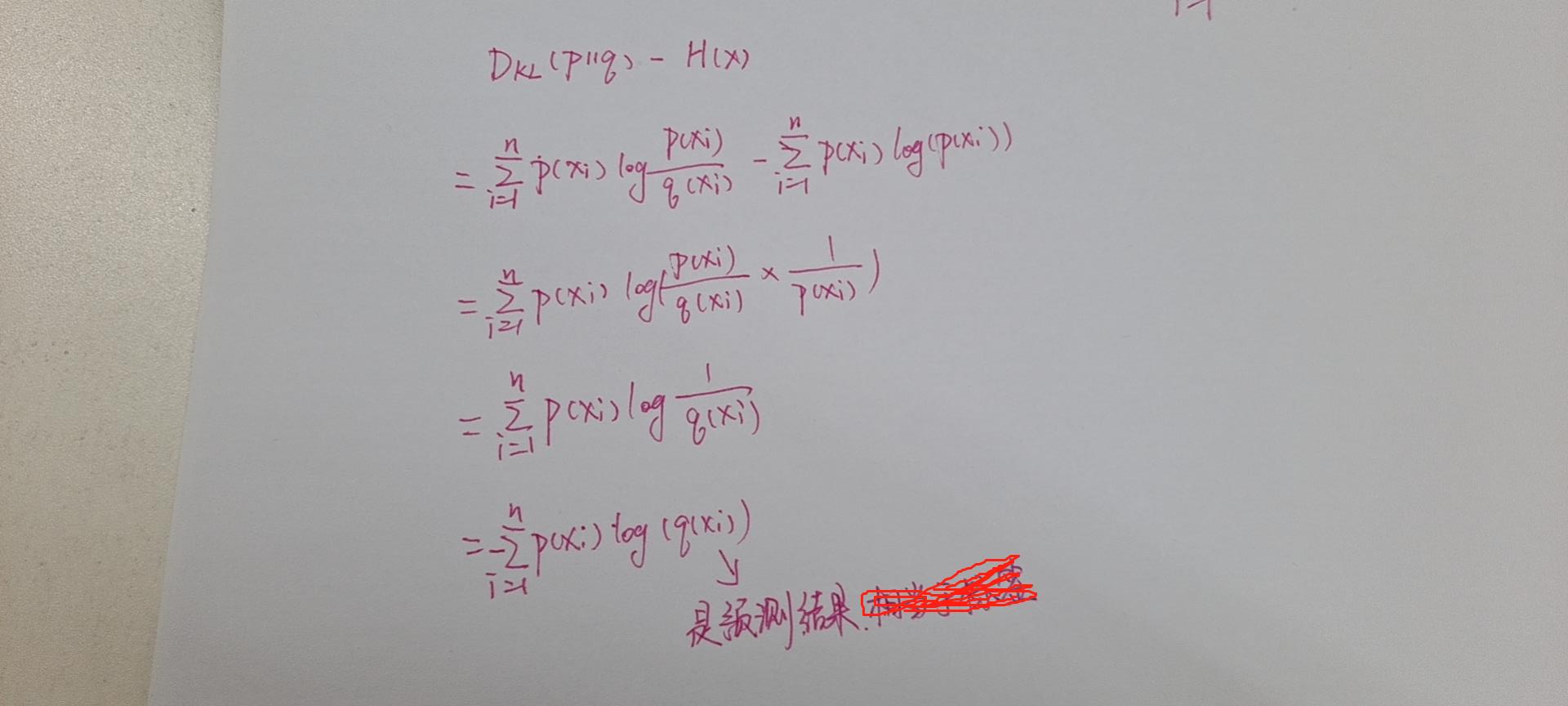
交叉熵=KL散度+信息熵

KL散度的公式：

交叉熵的公式：

对于二分类和多分类的交叉熵公式如下：这是因为在二分类中对用的网络最后一层是sigmoid，多分类中是softmax：

我们知道KL散度可以衡量两个分布之间的差异性，对于一个样本空间而言，其信息熵是确定的，因此也可以用交叉熵衡量两个分布的差异



# MPL的流程

数据：有标签数据：无标签数据=1：config. UDA\_DATA 详见Dataset.py

类别数目：10

训练过程(train.py)

Line82:合并后的数据经过teacher提取特征，每张图片输出10维的数据，可以理解为分别对应着10个类别

Line83:

UdaCrossEnroy.py:

Line21~24:将经过CNN的结果区分开，这里面的图片为有标签，无标签img，无标签img经过变换后的，

即为: logits['l'], logits['ori'], logits['aug']

Line27:有标签数据的label

有监督部分

Line31~35:计算了有标签图片的CE，这里没有经过softmax，所以在计算中每张图片对应的所有类别概率和并不一定为1

Line37:对有标签的结果取softmax，就相当于转化为概率

Line38:只输出对label类别的概率，也就是预测正确的概率

Line41~43:计算一个阈值l\_threshold，随着训练次数的增多，这个值会逐渐接近于1

Line45~47:建立一个mask，记录图片的正确概率值是否超过了阈值

Line48:求平均CE

无监督部分

Line51:计算无标签原图对应各个类别的概率，其中有个可调参数

Line55~57:计算无标签的CE，这里labels['ori']就相当于是概率，参考一下前文CE公式

Line59:计算无标签原图中预测概率最大值

Line61:判断这个概率值是否大于设定的阈值config.UDA\_THRESHOLD

Line62:存放结果，如果大于等于为1，否则为0

Line66~67:在用原图生成标签的时候，无法判断这个标签信息是否是正确的，所以规定只有概率大于阈值config.UDA\_THRESHOLD的标签我们才使用，也就是说只有这部分图片的信息计入loss，是有用的

Line92~98:我们希望student去学会teacher的知识，也就是student可以和teacher的结果接近，所以y\_true=tf.stop\_gradient(tf.nn.softmax(logits['aug'], -1))

即为

Line100~106:计算student在label图片上的损失，因为是带label的图片这里的标签是真实标签

Line118~121:根据cross\_entroy['s\_on\_u']反向传播，也就是让student逐步逼近teacher，更新后的student记为student’

Line123~128:用studen’再过一遍有标签图片，计算CE

Line130:计算新旧图片产生CE的区别。即为:dot\_product，新减旧

Line131~135: moving\_dot\_product是一个随机变量

dot\_product = 0.99\*dot\_product – 0.99\*moving\_dot\_product

为了给上一步的结果加入微小的随机性？！

Line147~149:对于teacher而言，一部分loss来自于teacher对带标签图片的判断；一部分来自于对无标签的判断，随着训练次数增加，网络效果越来越好，teacher对无标签loss的信任度也随之增加，体现在uda\_weight这个变量中；而student的效果好坏在带标签图片上得以验证，如果student效果好，更新后的student’产生的CE就比较小，则dot\_product大多数应该是负数，反之dot\_product就会比较大，这种差异是根据teacher对无标签变化后图片的预测得来的，也就是Line149（？！），以上组成了teacher的loss

Line150~154：teacher权重的更新

PS:teacher和student并不是同时开始训练的，相关设置可参考

LearningRate.py的Line54，

config.STUDENT\_LR\_WAIT\_STEPS

有时候权重是不更新的