**李宏毅av18603573笔记**

[1.SVM 1](#_Toc1338332290)

[2.GAN 1](#_Toc1827715866)

[1.预备知识 1](#_Toc1163556400)

[2.GAN的原理推导 2](#_Toc1015331010)

[2.1 求解D 2](#_Toc52038088)

[2.2 如何理解V(G,D)等价于P\_data和P\_G的分布差异 3](#_Toc790298846)

[2.3 求解G 4](#_Toc193356078)

[2.4 In practice 求解D和G 4](#_Toc1474524503)

# 1.SVM

pdf:

<http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Lecture/SVM%20(v5).pdf>

SVM的知识点主要是两个:**hinge loss**和**核函数trick**.

SVM的本质是给定线性可分训练数据集，通过**间隔最大化**求解相应的凸二次规划问题学习得到分离超平面:

IMG_256

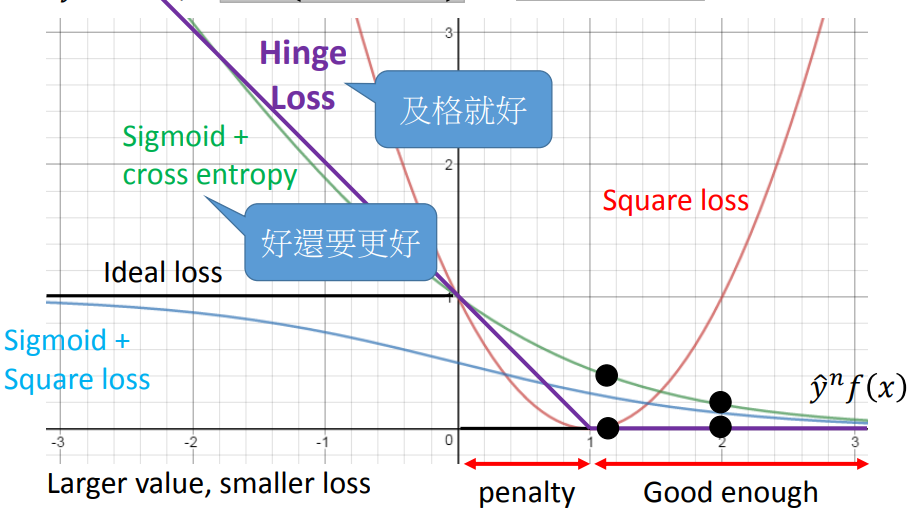
对线性可分得训练数据集而言，线性可分分离超平面有无穷多个，但是几个间隔最大的分离超平面是唯一的。（这里的间隔最大化也称为硬间隔最大，与训练数据集近似可分时软间隔最大化相对应）

## 1.hinge loss

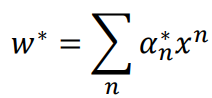
如下:

2018-03-22 19-12-52屏幕截图

对比softmax和svm loss,svm loss倾向于好一个margin就可以,而softmax倾向于一直好下去.



W的本质是Linear combination of data points



支撑向量的定义:

2018-03-22 22-00-36屏幕截图

## 2.核trick

对于**线性不可分**的数据集，可以利用**核函数（kernel）**将数据转换成易于分类器理解的形式,**从一个特征空间到另一个特征空间的映射**。在通常情况下，这种映射会**将低维特征空间映射到高维空间**。

SVM优化中一个特别好的地方就是，所有的运算都可以**写成内积（inner product）的形式**。向量的内积指的就是两个向量相乘，之后得到单个标量或者数值。我们可以**把内积运算替换成核函数**，而并不必做简化处理。

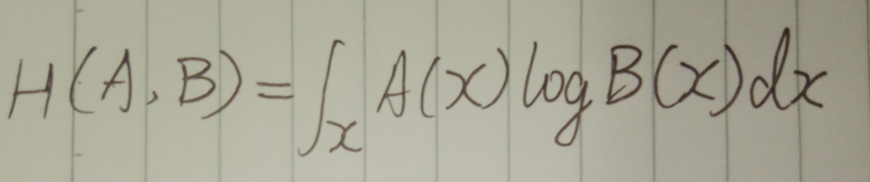
# 2.GAN

pdf :

<http://speech.ee.ntu.edu.tw/~tlkagk/courses/MLDS_2017/Lecture/GAN%20(v11).pdf>

## 1.预备知识

p真实分布,q模拟分布



信息熵: 完美编码.

H(p,p)

交叉熵: 不完美编码.用分布q模拟真实分布p所需的信息.

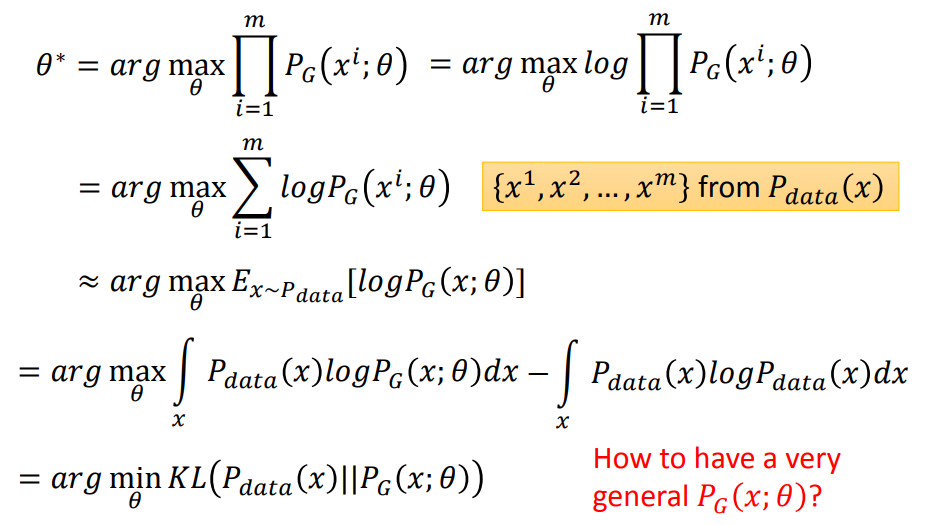
H(q,p)

相对熵: 即KL散度.表示用分布q模拟真实分布p相比用p模拟p，所需的额外信息。

KL(p||q)=H(p,p)-H(q,p)

寻找一个参数,可以最大化,H(p,q)-H(p,p)=-KL(p||q)

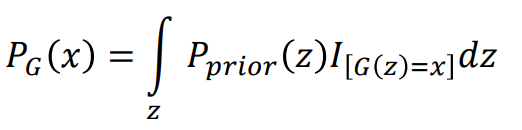
相当于寻找一个参数,最小化KL(p||q)



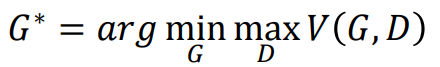
## 2.GAN的原理推导

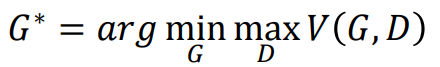
对于GAN的G,如果z服从某种分布,那么G(z)必然也是一个分布

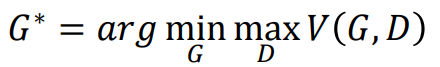
对于GAN生成的x,最简单的求解方法是遍历整个norm分布上所有的z:



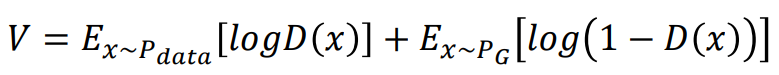
但是这种方法很难求极大似然,所以转化成下面这个:



求一个G,这个G使得的取值最小.

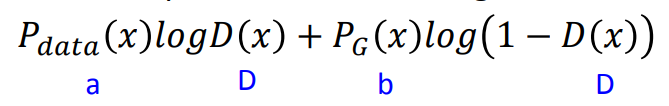
同时求一个D,在G固定的前提下使得的取值最大.

这里V(G,D)衡量的是P\_data和P\_G的分布差异.

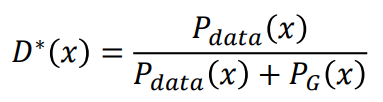


### 2.1 求解D

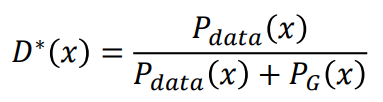
求解D时,我们认为G已经固定,这时最大化V就变成了最大化下式:

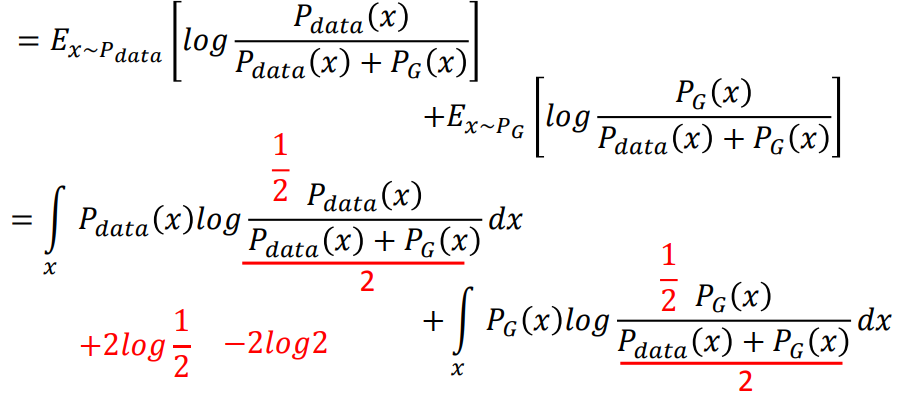


注意由于G固定,这里小写a和b都是scalar.将这个式子对D(x)求导,导数=0处就可以得到所求的D.



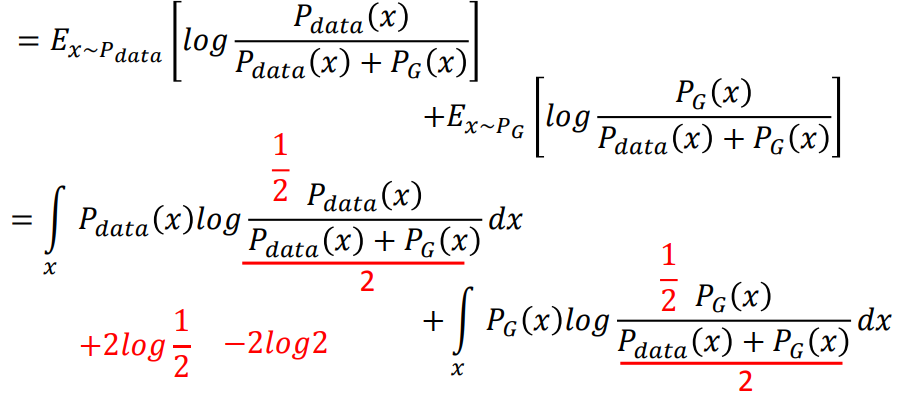
### 2.2 如何理解V(G,D)等价于P\_data和P\_G的分布差异

针对2.1中求出的,我们把它带入V(G,D),得到下式.



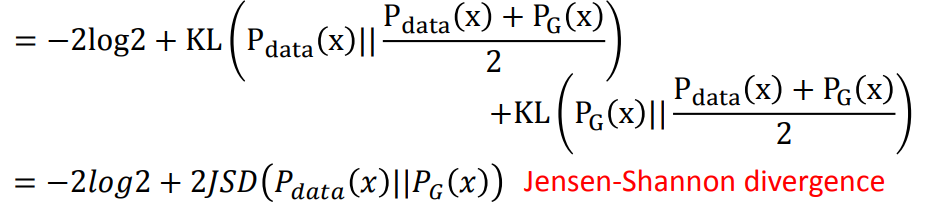
ps:

(1)一个trick,上下同时除以2

(2)期望用积分方式求解.

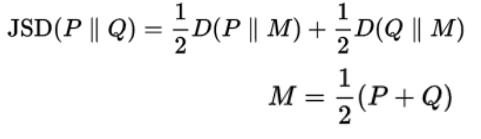
再简化就是下式,其实就是P\_data和P\_G的js散度.

说明,给定G时,寻找一个D让V最大,确实在衡量P\_data和P\_G的分布差异.



Ps:

JS散度公式,js散度是对称的,KL散度不对称:



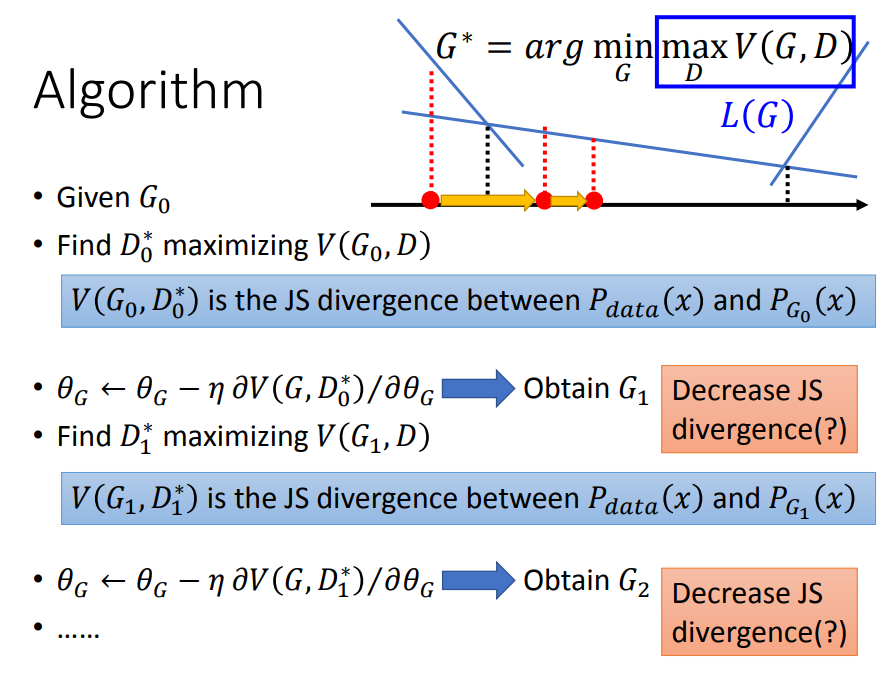
Js在[0,log2]之间,那么V(G,D)就在[-2log2,0]

### 2.3 求解G

最开始先随机一个G0,求D0,再通过导数=0的点求解G1,再求D1

......

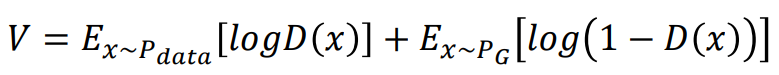
会发现这个流程其实就是不断地最小化P\_data和P\_Gi的JS散度



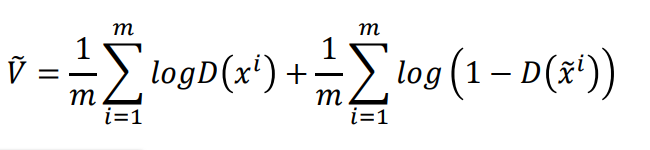
### 2.4 In practice 求解D和G

固定G的情况下,找一个D最大化真假数据的KL散度;找到D之后,再来找一个G最小化真假数据的JS散度.

(1)求解D



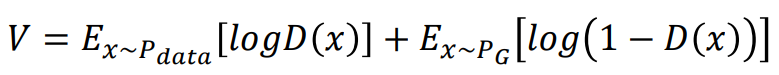
上式在实际中,先求解D,实际中不用积分,而是用samples的平均值.



所以求解D就是,找一些点,让他们最大化正负类的KL 散度(等同于最小化分类loss).

实际上应该多次训练D才能找到让KL最大的D

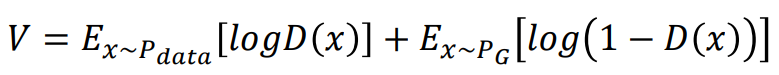
(2)求解G

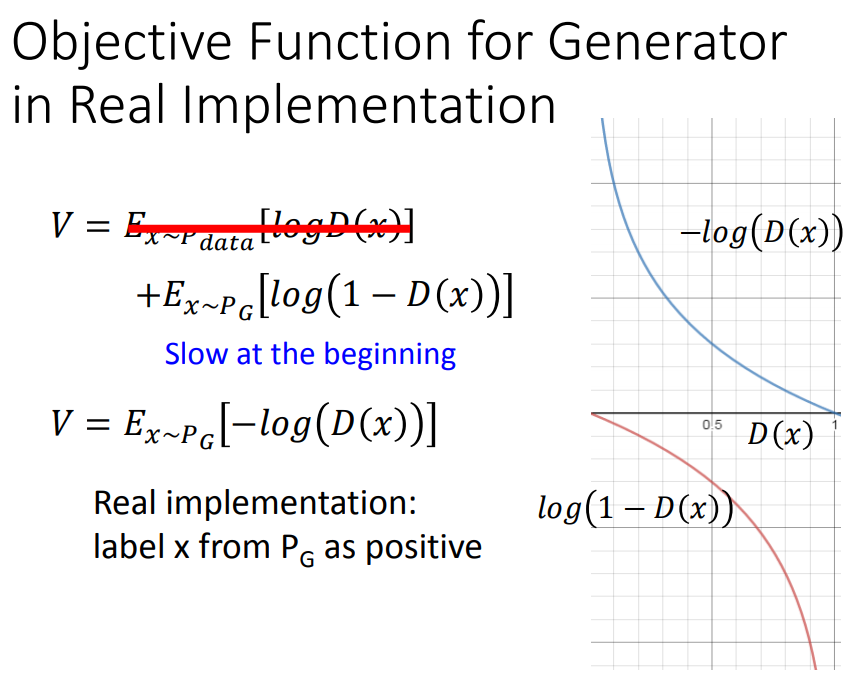


固定D的情况下,最小化上面的式子.

这个时候左半边不受G控制,只要最小化右半边就可以.

G只优化一次.

实际中,我们不去最小化,

而是最小化

参见函数,log(1-D(x) )在D(x)很小的时候,也就是G生成的图很难骗过D的时候,优化速度会非常慢.

