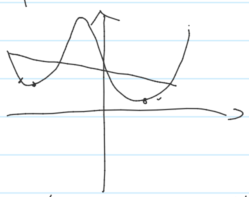
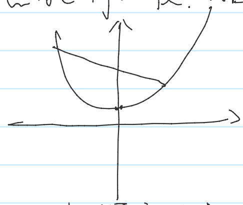
# 凸集

2020年3月27日 11:15

## 1. 凸集和凸函数.

我们先给出凸集的概念. 这个概念可以帮助我们解决最值和极值的问题
之前在求取最值时我们需要先找出所有的极值然后进行比较. 从图形上来看. 会有这样的情况.





第一种情况则有二个极值, 需要比较. 而第二种取得的极值就是最值. 我们如何描述这种区别呢?
可以直观的看到. 对同一 $x$ 值, 若曲线上的点都小于直线上的点, 那么这就是一个严格的下凸函数. 极值点就是最值点
我们用数学语言加以描述.

首先我们需要注意. 直线上的点也需要在定义域内. 对于横坐标不是实数点, 例如信息论中, 自变量是概率分布. 这是需要考虑的. 那么就有凸集的概念:

若对区域 $D$ 内任意两点 $\alpha$ 和 $\beta$, $\alpha\in D$, $\beta\in D$, 均有 $\lambda\alpha+(1-\lambda)\beta\in D$, $0\le\lambda\le 1$ 则称区域 $D$ 为凸域.

通俗一点讲. 若两点 $\alpha$ 和 $\beta$ 在凸域内. 则 $\alpha$ 和 $\beta$ 之间的线段也在整个区域 $D$ 内. 例如实数域是凸域, 但整数域和有理数域不是凸域. 在信息论中, 我们还要考虑概率分布组成的矢量集合. 例如

$$S_p=\vec{p},\quad p_n\ge 0,\quad n=1,2,\cdots,N,\quad \sum_{n=1}^{N}p_n=1.$$

这是一个 $N-1$ 维的凸域. 简单证明一下., 设 $\alpha=(p_{\alpha1},p_{\alpha2},\ldots,p_{\alpha N})$, $\beta=(p_{\beta1},\cdots,p_{\beta N})$, 则有.

$$\lambda\alpha+(1-\lambda)\beta=(\lambda p_{\alpha1}+(1-\lambda)p_{\beta1},\cdots,\lambda p_{\alpha N}+(1-\lambda)p_{\beta N})$$

$$\sum_{i=1}^{N}\lambda p_{\alpha i}+(1-\lambda)p_{\beta i}=\lambda+(1-\lambda)=1.$$

所以概率分布的集合也是凸集
接下来我们可以定义凸函数和凹函数以解决凸优化问题.

若在凸域 $D$ 上的 $f(x)$ 满足关系式

$$f(\lambda\alpha+(1-\lambda)\beta)\le\lambda f(\alpha)+(1-\lambda)f(\beta)\qquad \alpha,\beta\in D\qquad 0\le\lambda<1.$$

则称函数为下凸函数. 若上式中的不等式是严格不等式, 则称 $f(x)$ 是定义在凸域上的严格下凸函数.

若在凸域 $D$ 上的 $f(x)$ 满足关系式

$$f(\lambda\alpha+(1-\lambda)\beta)\ge\lambda f(\alpha)+(1-\lambda)f(\beta)\qquad \alpha,\beta\in D\qquad 0\le\lambda<1$$

则称函数为上凸函数. 若上式中的不等式是严格不等式. 则称 $f(x)$ 是定义在凸域上的严格上凸函数.

## 2. Jensen不等式

Jensen不等式是以丹麦数学家 Johan Jensen 命名的. 内容如下:
设 $f(x)$ 是凸域上的凸函数(下凸函数), 且 $a_m\in D$, $m=1,2,\cdots,M$. 则对 $0\le\lambda_m\le 1$, $\sum_{m=1}^{M}\lambda_m=1$ 有.

$$f\left(\sum_{m=1}^{M}\lambda_m a_m\right)\le\sum_{m=1}^{M}\lambda_m f(a_m)$$

这个公式和凸函数的定义是紧密相关的. 可以用数学归纳法证明.
当 $M=1$, $M=2$ 时. 依照定义成立.
当 $M=k$ 时, 若有

$$f\left(\sum_{m=1}^{k}\lambda_m a_m\right)\le\sum_{m=1}^{k}\lambda_m f(a_m)$$

当 $M=k+1$ 时. 则

$$f\left(\sum_{m=1}^{k}\lambda_m a_m+\lambda_{k+1}a_{k+1}\right)=f\left[\lambda_{k+1}a_{k+1}+(1-\lambda_{k+1})\sum_{m=1}^{k}\frac{\lambda_m}{1-\lambda_{k+1}}a_m\right]$$

由于 $\sum^{k}\frac{\lambda_m}{\ }=1$ 所以 $\sum^{k}\frac{\lambda_m}{\ }a_m$ 符合定义域. 因为假设了在 $M=k$ 时成立.

$$f\left(\sum_{m=1}^{k}\lambda_m a_m+\lambda_{k+1}a_{k+1}\right)=f\left[\lambda_{k+1}a_{k+1}+(1-\lambda_{k+1})\sum_{m=1}\frac{\lambda_m}{1-\lambda_{k+1}}a_m\right]$$

由于 $\sum_{m=1}^{k}\frac{\lambda_m}{1-\lambda_{k+1}}=1$，所以 $\sum_{m=1}^{k}\frac{\lambda_m}{1-\lambda_{k+1}}a_m$ 符合定义域．因为假设了在 $k$ 时成立．

由定义式可得

$$f\left(\sum_{m=1}^{k}\lambda_m a_m+\lambda_{k+1}a_{k+1}\right)\le\lambda_{k+1}f(a_{k+1})+(1-\lambda_{k+1})f\left(\sum_{m=1}^{k}\frac{\lambda_m}{1-\lambda_{k+1}}a_m\right)$$

且由假设．

$$f\left(\sum_{m=1}^{k}\frac{\lambda_m}{1-\lambda_{k+1}}a_m\right)\le\sum_{m=1}^{k}\frac{\lambda_m}{1-\lambda_{k+1}}f(a_m)$$

代入得．

$$f\left(\sum_{m=1}^{k+1}\lambda_m a_m\right)\le\sum_{m=1}^{k+1}\lambda_m f(a_m)$$

由此得证．由于 $\sum_{m=1}^{k+1}\lambda_m=1$．我们也可以认为整个过程是求取期望的过程．则有

$$f[E(x)]\le E[f(x)].$$

对于连续变量．

$$f\left[\int p(x)\cdot x\,dx\right]\le\int p(x)\cdot f(x)\,dx$$

最后我们探究一下等号的取得条件．从定义考虑．取等则会在极值点取得、向区间两端定义域压缩成一个点，在概率中也就是等概分布时取得等号．

## 3. 凸优化浅谈．

凸优化不仅在信号处理中应用广泛．在机器学习中应用也很广．虽然不是所有优化目标都是凸函数．但是利用凸优化进行求解仍然是一个很大的思路前提．例如梯度下降法．在信息论中，信息熵，互信息以及鉴别信息的优化问题都与凸优化有关．