Notes for Convex Functions

Xiaoyu Chen

这份笔记是在阅读 «Log-Concave Polynomial II» 时的一个衍生产物. 为了能够更好了的理解 strongly log-concave polynomial 为什么能够在采样问题中带来 rapix mixing, 所以我专门去学习了一下 Convex functions. 主要是想在 concave 和 Hessian matrix 的半负定之间建立起比较严谨的联系这篇文章的主要内容来自于知乎大佬的博客

1 定义

在本文中, 我们谈论的是 convex funcion (凸函数), concave function (凹函数) 其实区别不大, 只是方向不一样而已.

Define 1 (convex funcion)**.** *对于 $f: \mathbb{R}^n \to \mathbb{R}$, 如果对于 f 定义域中的任意两个自变量 x_1, x_2 和任意 $\theta \in [0,1]$, 都有*

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$$

*则说, f 是***凸函数***. (凹函数直接改变不等号的方向即可)*

我们希望有方便的工具来判断一个函数是否是凸函数, 这就需要找到一些充要条件. 下面给出两个重要的判别条件, 我们比较关心的是后一个, 不过我目前找到的资料中, 这两个判据的证明都是有依赖关系的.

2 一阶判据

Theorem 1 (First Order Condition)**.** *函数 f 是凸函数, 当且仅当, f 的定义域上的任意两点 x_1, x_2, 都有:*

$$f(x_1) \geq f(x_2) + \nabla f(x_2)(x_1 - x_2)$$

△ 注: 这里 $x_1 - x_2$ 是向量相减, 图形上看, 是一个从点 x_2 指向点 x_1 的一个向量.

Remark 1 (几何意义)**.** *上面的不等式中, 如果 x_1 是一个自由变量, 则 $f(x_2) + \nabla f(x_2)(x_1 - x_2)$ 描述了一个 $n+1$ 维空间中的超平面. 这个不等式的意思就是***整个函数 f 都在这个超平面上方***. 这一点还是比较直观的.*

proof. 凸函数 $\Rightarrow f(x_1) \geq f(x_2) + \nabla f(x_2)(x_1 - x_2), \forall x_1, x_2 \in \mathrm{dom}\, f$:

$$
\begin{aligned}
f(\theta x_1+(1-\theta)x_2) &\le \theta f(x_1)+(1-\theta)f(x_2)\\
f(x_2+\theta(x_1-x_2)) &\le f(x_2)+\theta(f(x_1)-f(x_2))\\
f(x_2+\theta(x_1-x_2))-f(x_2) &\le \theta(f(x_1)-f(x_2))\\
\frac{f(x_2+\theta(x_1-x_2))-f(x_2)}{\theta} &\le f(x_1)-f(x_2)\\
f(x_1) &\ge f(x_2)+\frac{f(x_2+\theta(x_1-x_2))-f(x_2)}{\theta}
\end{aligned}
$$

因为这个式子需要对所有的 x_1, x_2, θ 成立, 我们不妨让 θ 逼近 0. 来考虑下面的极限:

$$
\lim_{\theta\to 0}\frac{f(x_2+\theta(x_1-x_2))-f(x_2)}{\theta}
$$

令 $g(\theta) := f(x_2+\theta(x_1-x_2))$ (也就是说 $x(\theta)=x_2+\theta(x_1-x_2), g(\theta)=f(x(\theta))$), 则上述极限相当于在对 g 求导.

$$
\begin{aligned}
\lim_{\theta\to 0}\frac{f(x_2+\theta(x_1-x_2))-f(x_2)}{\theta}=\frac{g(\theta)-g(0)}{\theta}=\left.\frac{\mathrm{d}g}{\mathrm{d}\theta}\right|_{\theta=0}=\frac{\mathrm{d}f}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}\theta}&=(x_1-x_2)\left.\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{x=x_2}\\
&=\nabla f(x_2)(x_1-x_2)
\end{aligned}
$$

△ 注: 其实, $\dfrac{\mathrm{d}f(x_2)}{\mathrm{d}x_2}=\nabla f(x_2)$ 这一步是比较迷的. 不过从全微分的角度, 还是能够解释的, 期待以后补完这个东西的详细定义.

所以, 和上面结合起来就是:

$$
f(x_1)\ge f(x_2)+\nabla f(x_2)(x_1-x_2)
$$

下面来说明: $f(x_1)\ge f(x_2)+\nabla f(x_2)(x_1-x_2)\Rightarrow$ 凸函数:

因为左边的不等式对于任何的 x_1, x_2 都是适用的, 所以对于任意的 θ, 令 $x=\theta x_1+(1-\theta)x_2$, 我们有如下结论成立:

$$
\begin{cases}
f(x_1)\ge f(x)+\nabla f(x)(x_1-x)\\
f(x_2)\ge f(x)+\nabla f(x)(x_2-x)
\end{cases}
$$

把这两个式子按照 x_1, x_2 的比例相加, 就有:

$$
\begin{aligned}
&\theta f(x_1)+(1-\theta)f(x_2)\ge f(x)+\nabla f(x)(\theta x_1+(1-\theta)x_2)-\nabla f(x)x\\
&\theta f(x_1)+(1-\theta)f(x_2)\ge f(x)=f(\theta x_1+(1-\theta)x_2)
\end{aligned}
$$

□

3 二阶判据

Theorem 2 (Second-Order Condition). *函数 f 为凸函数, 当且仅当对于任意自变量 x_0, 其 Hessian 矩阵 $H(x_0)$ 都是半正定的.*

proof. 凸函数 $\Rightarrow \forall x_0, H(x_0)$ 半正定:

f 在 x_0 点的泰勒展开式 (Peano 余项) 如下:

$$
f(x)=f(x_0)+\nabla f(x_0)(x-x_0)+\frac{1}{2}(x-x_0)^T H(x_0)(x-x_0)+o\left((x-x_0)^T(x-x_0)\right)
$$

我们令 $x = x_0 + d$ (d 为一个微小的向量), 可以将上面的式子改写成下面这样:

$$f\left(x_0+d\right)=f\left(x_0\right)+\nabla f\left(x_0\right) d+\frac{1}{2} d^T H\left(x_0\right) d+o\left(d^T d\right)$$

因为 f 是凸函数, 所以由第一判据

$$f(x_0+d) \geq f(x_0)+\nabla f(x_0) d$$

上面两个不等式联合起来可以发现:

$$\begin{aligned}
\frac{1}{2} d^T H\left(x_0\right) d+o\left(d^T d\right) & \geq 0 \\
\lim _{||d||_2 \rightarrow 0} \frac{1}{2} \frac{d^T H\left(x_0\right) d}{d^T d}+\frac{o\left(d^T d\right)}{d^T d} & \geq 0 \\
\lim _{||d||_2 \rightarrow 0} \frac{d^T H\left(x_0\right) d}{d^T d} & \geq 0
\end{aligned}$$

所以 $H(x_0)$ 半正定.

$\forall x_0, H(x_0)$ 半正定 $\Rightarrow$ 凸函数:

f 在 x_0 点的泰勒展开式 (Lagrange 余项余项) 如下

$$f(x)=f\left(x_0\right)+\nabla f\left(x_0\right)\left(x-x_0\right)+\frac{1}{2}\left(x-x_0\right)^T H(\xi)\left(x-x_0\right)$$

因为 $H(\xi)$ 正定, 所以我们可以立即判断出

$$f(x) \geq f(x_0)+\nabla f(x_0)(x-x_0), \forall x, x_0 \qquad \square$$

不难发现, concave function 的 Hessian 阵一定是负定的.