Convexity

Divya Bhagia

Boston College

August 20, 2019

Overview

Overview

Convex Sets and Convex Hull

Definition

In real vector space V, a set $S \subset V$ is a **convex set** if

$$\lambda x + (1 - \lambda) y \in S$$

for any $\lambda \in [0, 1]$ and $x, y \in S$.

Definition

In a vector space V, the **convex hull** of set $S \subset V$, denoted by $Co(S)$ is the intersection of all convex sets in V that contain S. $Co(S)$ can be interpreted as the smallest convex set that covers S.

Convex Sets

Proposition

In a vector space V, the set $S \subset V$ is convex *iff* any convex combination of $x_1, x_2, ..., x_n \in S$ given by the vector $\sum_{i=1}^{n} \lambda_i x_i$ is also in S where $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{R}_+$ and $\sum_{i=1}^{n} \lambda_i = 1$.

Proof: $\Leftarrow$ is trivial.
$\Rightarrow$: Proof by induction. If $n = 1$, the statement is trivial. If $n = 2$, then $\lambda_1 x_1 + \lambda_2 x_2 \in S$ by definition of convexity. Now assume statement is true for $n = k$. With $n = k + 1$, consider any $\lambda_1, \lambda_2, ..., \lambda_{k+1} \in \mathbb{R}_+$ with $\sum_{i=1}^{k+1} \lambda_i = 1$, then we have

$$\sum_{i=1}^{k+1} \lambda_i x_i = \sum_{i=1}^{k} \lambda_i x_i + \lambda_{k+1} x_{k+1}$$
$$= \left(\sum_{j=1}^{k} \lambda_j\right) \sum_{i=1}^{k} \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j} x_i + \lambda_{k+1} x_{k+1} \qquad \square$$

Overview

Hyperplanes

Definition

Given $p \in \mathbb{R}^n$ with $p \neq 0$, and $c \in \mathbb{R}$, the **hyperplane** generated by p and c is the set

$$H_{p,c} := \{z \in \mathbb{R}^n : p^T z = c\}$$

The sets $\{z \in \mathbb{R}^n : p^T z \geq c\}$ and $\{z \in \mathbb{R}^n : p^T z \leq c\}$ are called **half-space above** and **half-space below** the hyperplane $H_{p,c}$, respectively.

Separating Hyperplane Theorems

Minkowski's Theorem

Suppose that the convex sets $A, B \subset \mathbb{R}^n$ are disjoint (i.e. $A \cap B = \phi$). Then there exists $p \in \mathbb{R}^n$ with $p \neq 0$ and a value $c \in \mathbb{R}$ such that $p^T x \geq c$ for every $x \in A$ and $p^T y \leq c$ for every $y \in B$.

Separating Hyperplane Theorems

Weaker Version with Strict Separation

Suppose that the *closed* convex sets $A, B \subset \mathbb{R}^n$ are disjoint with at least one of them *bounded*. Then there exists $p \in \mathbb{R}^n$ with $p \neq 0$ and a value $c \in \mathbb{R}$ such that $p^T x > c$ for every $x \in A$ and $p^T y < c$ for every $y \in B$.

Proof: Define $d(A, B) = inf\{d_2(a, b) : a \in A, b \in B\}$. Let $\hat{a}$ and $\hat{b}$ be points that achieve it. Now let

$$p = \hat{a} - \hat{b} \qquad \text{and} \qquad c = \frac{||\hat{a}||^2 - ||\hat{b}||^2}{2}$$

Claim 1: $p^T x > c$ for every $x \in A$.
Proof by contradiction, assume that there exists $a \in A$ s.t. $p^T a \leq c$. This means that

$$(\hat{a} - \hat{b})^T a \leq \frac{||\hat{a}||^2 - ||\hat{b}||^2}{2} \tag{1}$$

Proof Cont.

Define $g(x) = ||x - \hat{b}||^2$, then $\nabla g(x) = 2(x - \hat{b})$.

$$\begin{aligned}
\nabla g(\hat{a})(a - \hat{a}) &= 2(\hat{a} - \hat{b})^T (a - \hat{a}) \\
&= 2(-||\hat{a}|| + (\hat{a} - \hat{b})^T a + \hat{b}^T \hat{a}) \\
&\leq 2\left(-||\hat{a}|| + \frac{||\hat{a}||^2 - ||\hat{b}||^2}{2} + \hat{b}^T \hat{a}\right) && \text{From (1)} \\
&= -||d - c||^2 < 0
\end{aligned}$$

This implies that $\exists \bar{\alpha} > 0$ s.t. $\forall \alpha \in (0, \bar{\alpha})$

$$g(\hat{a} + \alpha(\hat{a} - a)) < g(\hat{a})$$

But this contradicts that $\hat{a}$ was the closest point to $\hat{b}$. $\square$

Overview

Concave and Convex Functions

Definitions

Consider a function $f : A \to \mathbb{R}$, where A is a convex set in vector space V.

1. The function f is a **concave** function if

$$f(\lambda x' + (1-\lambda)x) \geq \lambda f(x') + (1-\lambda)f(x)$$

   for all x and $x' \in A$ and all $\lambda \in [0,1]$. If inequality is strict for all $x' \neq x$ and all $\lambda \in (0,1)$, then we say that the function is **strictly concave**.

2. The function f is a **convex** function if

$$f(\lambda x' + (1-\lambda)x) \leq \lambda f(x') + (1-\lambda)f(x)$$

   for all x and $x' \in A$ and all $\lambda \in [0,1]$. If inequality is strict for all $x' \neq x$ and all $\lambda \in (0,1)$, then we say that the function is **strictly convex**.

First-Order Characterization of Concavity

Theorem

Suppose the function $f : S \to \mathbb{R}$ is a C^2 function, where S is a convex and open set in $\mathbb{R}^n$. Then f is concave in S *iff*

$$f(x) - f(x^0) \leq \nabla f(x^0)^T (x - x^0) = \sum_{i=1}^{n} \frac{\partial f(x^0)}{\partial x_i}(x_i - x_i^0)$$

for all $x, x^0 \in S$.

Proof: $\Rightarrow$: Let $x, x^0 \in S$. Since f is concave,

$$\lambda f(x) + (1-\lambda) f(x^0) \leq f(\lambda x + (1-\lambda)x^0)$$

for all $\lambda \in [0,1]$. Rearranging the above inequality, for all $\lambda \in (0,1]$, we obtain

$$f(x) - f(x^0) \leq \frac{f(x^0 + \lambda(x - x^0)) - f(x^0)}{\lambda} \qquad (\#)$$

For $\lambda \to 0$, the RHS of $(\#)$ approaches $\nabla f(x^0)^T (x - x^0)$.

First-Order Characterization of Concavity

Proof Cont.

$\Leftarrow$: Define $z = \lambda x + (1-\lambda)x^0$. Notice that $z \in S$ because S is convex. Then given the statement of the theorem,

$$f(x) - f(z) \leq \nabla f(z)^T (x - z) \quad (1)$$
$$f(x^0) - f(z) \leq \nabla f(z)^T (x^0 - z) \quad (2)$$

Multiplying (1) by $\lambda > 0$ and (2) by $1 - \lambda > 0$, we obtain

$$\lambda(f(x) - f(z)) + (1-\lambda)(f(x^0) - f(z)) \leq \nabla f(z)^T \underbrace{[\lambda(x - z) + (1-\lambda)(x^0 - z))]}_{\lambda x + (1-\lambda)x^0 - z = 0}$$

Then re-arranging the above equation gives us

$$\lambda f(x) + (1-\lambda) f(x^0) \leq f(z) = f(\lambda x + (1-\lambda)x^0)$$

Thus f is concave. $\square$

First-Order Characterization of Strict Concavity

Corollary

f is strictly concave *iff* the inequality in the previous theorem is strict when $x \neq x^0$.

Proof: $\Rightarrow$: Suppose that f is strictly concave in S. Then, inequality ($\#$) is strict for all $\lambda \in (0,1)$ and all $x \neq x^0$.
$\Leftarrow$: With $z = x^0 + \lambda(x - x^0)$, we have

$$f(x) - f(x^0) < \frac{f(z) - f(x^0)}{\lambda} \leq \frac{\nabla f(x^0)^T (z - x^0)}{\lambda} = \nabla f(x^0)^T (x - x^0) \qquad \square$$

First-Order Characterization of Convexity

Theorem

Suppose the function $f : S \to \mathbb{R}$ is a C^1 function, where S is a convex and open set in $\mathbb{R}^n$. Then f is convex in S *iff*

$$f(x) \geq f(x^0) + \nabla f(x^0)^T(x - x^0) = \sum_{i=1}^{n} \frac{\partial f(x^0)}{\partial x_i}(x_i - x_i^0)$$

for all $x, x^0 \in S$.

In addition, f is strictly convex *iff* the inequality is strict when $x \neq x^0$.

Second-Order Characterization in $\mathbb{R}$

Theorem

Suppose the function $f : S \to \mathbb{R}$ is a C^2 function, where $S \subset \mathbb{R}$ is convex and open.

1. f is convex *iff* $f''(x) \geq 0$ for any $x \in S$.
2. f is concave *iff* $f''(x) \leq 0$ for any $x \in S$.

Replace with strict inequalities to obtain equivalent statements for strict convexity and strict concavity.

Proof: We do a mean-value expansion of the function around x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x^*)(x - x_0)^2$$

where x^* lies between x_0 and x. By our hypothesis, $f''(x^*) \geq 0$. Let $x_0 = \lambda x_1 + (1 - \lambda)x_2$ and $x = x_1$, then

$$f(x_1) \geq f(x_0) + f'(x_0)(1 - \lambda)(x_1 - x_2) \tag{1}$$

Similarly taking $x = x_2$

$$f(x_2) \geq f(x_0) + f'(x_0)\lambda(x_1 - x_2) \tag{2}$$

Multiplying (1) by λ and (2) by $(1 - \lambda)$ and adding finishes the proof for (1). Rest of the statements follow easily. □

Second-Order Characterization of Convexity and Concavity

Theorem

Suppose the function $f : S \to \mathbb{R}$ is a C^2 function, where S is a convex and open set in $\mathbb{R}^n$.

1. f is convex *iff* it's Hessian matrix $D^2 f(x)$ is positive semi-definite for any $x \in S$.
2. f is concave *iff* it's Hessian matrix $D^2 f(x)$ is negative semi-definite for any $x \in S$.

Proof: $\Leftarrow$: Take two points $x, x^0 \in S$ and let $t \in [0, 1]$. Define

$$g(t) = f(x^0 + t(x - x^0)) = f(tx + (1 - t)x^0)$$

Differentiating with respect to t

$$g'(t) = \sum_{i=1}^{n} f_i(x^0 + t(x - x^0))(x_i - x_i^0) = (x - x^0)^T [\nabla f(x^0 + t(x - x^0)]$$

Second-Order Characterization of Convexity and Concavity

Proof Cont.

Differentiating again with respect to t

$$g''(t) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(x^0 + t(x - x^0))(x_i - x_i^0)(x_j - x_j^0)$$
$$= (x - x^0)^T [D^2 f((x^0 + t(x - x^0)](x - x^0)$$

Because $D^2 f$ is positive semi-definite, therefore $g''(t) \geq 0$ for any $t \in [0, 1]$ which implies that g is convex. Then it follows

$$g(t) = g(t.1 + (1 - t).0) \geq tg(1) + (1 - t)g(0) = tf(x) + (1 - t)f(x^0)$$

and f is also convex.

$\Rightarrow$: Suppose f is convex. Take $x \in S$ and arbitrary vector $h = (h_1, ..., h_n)$, then there exists $a > 0$ s.t. $x + th \in S$ $\forall t$ with $|t| < a$. Let $I = (-a, a)$.

Second-Order Characterization of Convexity and Concavity

Proof Cont.

Define the function p on I by $p(t) = f(x + th)$. Since $p(.)$ is convex in I

$$p''(t) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(x + th)h_i h_j \geq 0$$

Plugging in $t = 0$, it follows that

$$h^T D^2 f(x) h \geq 0$$

for any $x \in S$ and any h, therefore $D^2 f(x)$ is positive semi-definite. Proof for (2) follows by setting $f = -f$. $\square$

Second-Order Partial Characterization of Strict Convexity and Concavity

Theorem

Suppose the function $f : S \to \mathbb{R}$ is a C^2 function, where S is a convex and open set in $\mathbb{R}^n$.

1. f is strictly convex **if** it's Hessian matrix $D^2 f(x)$ is positive definite for any $x \in S$.
2. f is strictly concave **if** it's Hessian matrix $D^2 f(x)$ is negative definite for any $x \in S$.

Proof: Define $g(.)$ as in the proof of the previous theorem. If $D^2 f(x)$ is positive definite, then for $x \neq x^0$, $g''(t) > 0$ for all $t \in [0, 1]$ which implies that $g(.)$ is strictly convex. Then we have

$$g(t) = g(t.1 + (1-t).0) > tg(1) + (1-t)g(0) = tf(x) + (1-t)f(x^0)$$

for all $t \in (0, 1)$. Proof for (2) follows by setting $f = -f$. $\square$

Overview

Quasiconcave and Quasiconvex Functions

Definitions

Conside a function $f : S \to \mathbb{R}$, where S is a convex set in vector space V.

1. The function f is a **quasiconcave** function if

$$f(\lambda x + (1-\lambda)x') \geq min\{f(x), f(x')\}$$

   for any $x, x' \in S$ and $\lambda \in [0, 1]$.

2. The function f is a **quasiconvex** function if

$$f(\lambda x + (1-\lambda)x') \leq max\{f(x), f(x')\}$$

   for any $x, x' \in S$ and $\lambda \in [0, 1]$.

Strictly Quasiconcave and Strictly Quasiconvex Functions

Definitions

Consider a function $f : S \to \mathbb{R}$, where S is a convex set in vector space V.

1. The function f is a **strictly quasiconcave** function if
$$f(\lambda x + (1-\lambda)x') > min\{f(x), f(x')\}$$
for any $x, x' \in S$ with $x \neq x'$ and $\lambda \in (0,1)$.
2. The function f is a **strictly quasiconvex** function if
$$f(\lambda x + (1-\lambda)x') < max\{f(x), f(x')\}$$
for any $x, x' \in S$ with $x \neq x'$ and $\lambda \in (0,1)$.

Quasiconcave and Quasiconvex Functions

Theorem

Consider a function $f : S \to \mathbb{R}$, where S is a convex set in vector space V.

1. f is quasiconcave *iff* it's upper contour set $C^a = \{x \in S : f(x) \geq a\}$ is convex in V for any $a \in \mathbb{R}$.
2. f is quasiconvex *iff* it's lower contour set $C_a = \{x \in S : f(x) \leq a\}$ is convex in V for any $a \in \mathbb{R}$.

Proof: $\Rightarrow$: Take any $a \in \mathbb{R}$ and $x, x' \in C^a$, then $f(x) \geq a$ and $f(x') \geq a$. Because f is quasiconcave, for any $\lambda \in (0, 1)$,

$$f(\lambda x + (1-\lambda)x') \geq min\{f(x), f(x')\}$$

Thus $\lambda x + (1-\lambda)x' \in C^a$ implying that C^a is convex.
$\Leftarrow$: Take $x, x' \in S$ and $\lambda \in [0, 1]$. Define $a = min\{f(x), f(x')\}$. Then $x, x' \in C^a = \{x \in S : f(x) \geq a\}$ and therefore $\lambda x + (1-\lambda)x' \in C^a$. This implies that $f(\lambda x + (1-\lambda)x') \geq min\{f(x), f(x')\}$. Proof for (2) is symmetric. $\square$

Quasiconcavity is Preserved under Positive Monotone Transformation

Theorem

Consider a function $f : S \to \mathbb{R}$, where S is a convex set in vector space V. If f is quasiconcave (quasiconvex) and $\phi : \mathbb{R} \to \mathbb{R}$ is weakly increasing then $\phi \circ f$ is quasiconcave (quasiconvex).

Proof: Take any $x, x' \in S$ and $\lambda \in [0,1]$, then

$$f(\lambda x + (1-\lambda)x') \geq min\{f(x), f(x')\}$$

Since ϕ is strictly increasing

$$\phi(f(\lambda x + (1-\lambda)x')) \geq \phi(min\{f(x), f(x')\}) = min\{\phi(f(x), \phi(f(x'))\}$$

and hence $\phi \circ f$ is quasiconcave. Argument for quasiconvexity follows symmetrically. $\square$.

First-Order Characterization of Quasiconcavity

Theorem

Let $f : S \to \mathbb{R}$ be a C^1 function, where S is a convex and open set in $\mathbb{R}^n$. Then f is quasiconcave on S *iff* for all $x, x^0 \in S$

$$f(x) \geq f(x^0) \Rightarrow \nabla f(x^0)^T(x - x^0) = \sum_{i=1}^{n} \frac{\partial f(x^0)}{dx_i}(x_i - x_i^0) \geq 0$$

Second-Order Characterization of Quasiconcavity

Theorem

Let $f : S \to \mathbb{R}$ be a C^2 function, where S is a convex and open set in $\mathbb{R}^n$. Then f is quasiconcave on S *iff* for every $x \in S$, the Hessian matrix $D^2 f(x)$ is negative semidefinite in subspace $\{z \in \mathbb{R}^n : \nabla f(x)^T z = 0\}$. That is

$$z^T D^2 f(x) z \leq 0 \quad \text{whenever} \quad \nabla f(x)^T z = 0$$

for every $x \in S$. If the Hessian matrix $D^2 f(x)$ is negative definite in the subspace $\{z \in \mathbb{R}^n : \nabla f(x)^T z = 0\}$ for every $x \in S$, then f is **strictly** quasiconcave.

Characterization through Bordered Hessian

Theorem

Let $f : S \to \mathbb{R}$ be a C^2 function, where S is a convex and open set in $\mathbb{R}^n$. Define the bordered Hessian determinants $B_r(x)$ as follows for each $r = 2, ..., n$:

$$B_r(x) = \begin{vmatrix} 0 & f_1(x) & \dots & f_r(x) \\ f_1(x) & f_{11}(x) & \dots & f_{1r}(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_r(x) & f_{r1}(x) & \dots & f_{rr}(x) \end{vmatrix}$$

1. A necessary condition for f to be quasiconcave is that $(-1)^r B_r(x) \geq 0$ for all $x \in S$ and all $r = 1, ..., n$.
2. A sufficient condition for f to be strictly quasiconcave is that $(-1)^r B_r(x) > 0$ for all $x \in S$ and all $r = 1, ..., n$.

Review

Review: Hessian Matrix

Definition

Suppose the function $f : S \to \mathbb{R}$ is a C^2 function, where S is a convex and open set in $\mathbb{R}^n$. The matrix

$$D^2 f(x) = (f_{ij}(x))_{n \times n}$$

is called the **Hessian** matrix of f at x where $f_{ij}(x) = \partial^2 f(x)/\partial x_i \partial x_j$.

Review: Quadratic Forms and Definiteness

Definition

Any $N \times N$ matrix M determines a quadratic form Q

$$Q(x) = Q(x_1, ..., x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = x^T M x$$

Definitions

1. The $N \times N$ matrix M is **negative semi-definite** if $x^T M x \leq 0$ for all $x \in \mathbb{R}^N$.
2. The $N \times N$ matrix M is **negative definite** if $x^T M x < 0$ for all $z \neq 0$ with $x \in \mathbb{R}^N$.
3. The $N \times N$ matrix M is **positive semi-definite** if $x^T M x \geq 0$ for all $x \in \mathbb{R}^N$.
4. The $N \times N$ matrix M is **positive definite** if $x^T M x > 0$ for all $z \neq 0$ with $x \in \mathbb{R}^N$.

Review: Definiteness of Matrices

Theorem

1. The $N \times N$ matrix M is **positive definite** if and only if all it's N leading principal minors are strictly positive.

2. The $N \times N$ matrix M is **negative definite** if and only if its N leading minors alternate in sigh as follows:

$$|M_1| < 0 \quad |M_2| > 0 \quad |M_3| < 0 \quad \ldots$$

The kth order leading principal minor should have the same sign as $(-1)^k$.

Review: Definiteness of Matrices

Theorem

1. The $N \times N$ matrix M is **positive semi-definite** if and only if every principal minor of A is ≥ 0.

2. M is **negative semi-definite** if and only if every principal minor of odd order is ≤ 0 and every principal minor of even order is ≥ 0.

Review: Definiteness of Matrices

Easy to show previous two theorems for $n = 2$. In which case the quadratic form is

$$Q(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

After some manipulation, we obtain

$$Q(x_1, x_2) = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \left(a_{22} - \frac{a_{12}^2}{a_{11}}\right)x_2^2$$

Thus we obtain

- $Q(x_1, x_2) > 0$ *iff* $a_{11} > 0$ and $a_{11}a_{22} - a_{12}^2 > 0$
- $Q(x_1, x_2) < 0$ *iff* $a_{11} < 0$ and $a_{11}a_{22} - a_{12}^2 < 0$