

Vũ Hữu Tiệp

Convex Optimization và Support Vector Machines

Theo blog: <http://machinelearningcoban.com>

(Đang trong quá trình xây dựng)

First Edition

June 24, 2017

Contents

16 Convex sets và convex functions	1
Index	19

Convex sets và convex functions

16.1 Giới thiệu

Từ đầu đến giờ, chúng ta đã làm quen với rất nhiều bài toán tối ưu. Học Machine Learning là phải học Toán Tối Ưu, và để hiểu hơn về Toán Tối Ưu, với tôi cách tốt nhất là tìm hiểu các thuật toán Machine Learning. Cho tới lúc này, những bài toán tối ưu các bạn đã nhìn thấy trong blog đều là các bài toán tối ưu không ràng buộc (unconstrained optimization problems), tức tối ưu hàm mất mát mà không có điều kiện ràng buộc (constraints) nào về nghiệm cả.

Không chỉ trong Machine Learning, trên thực tế các bài toán tối ưu thường có rất nhiều ràng buộc khác nhau. Ví dụ:

- Tôi muốn thuê một ngôi nhà cách trung tâm Hà Nội không quá 5km với giá càng thấp càng tốt. Trong bài toán này, giá thuê nhà chính là hàm mất mát (*loss function*, đôi khi người ta cũng dùng *cost function* để chỉ hàm số cần tối ưu), điều kiện khoảng cách không quá 5km chính là ràng buộc (constraint).
- Quay lại [bài toán dự đoán giá nhà theo Linear Regression](#), giá nhà là một hàm tuyến tính của diện tích, số phòng ngủ và khoảng cách tới trung tâm. Rõ ràng, khi làm bài toán này, ta dự đoán rằng giá nhà tăng theo diện tích và số phòng ngủ, giảm theo khoảng cách. Vậy nên một nghiệm được gọi là *có lý một chút* nếu hệ số tương ứng với diện tích và số phòng ngủ là dương, hệ số tương ứng với khoảng cách là âm. Để tránh các nghiệm ngoại lai không mong muốn, khi giải bài toán tối ưu, ta nên cho thêm các điều kiện ràng buộc này.

Trong Tối Ưu, một bài toán có ràng buộc thường được viết dưới dạng:

$$\begin{aligned} \mathbf{x}^* &= \arg \min_{\mathbf{x}} f_0(\mathbf{x}) \\ \text{subject to: } & f_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, p \end{aligned}$$

Trong đó, vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ được gọi là *biến tối ưu* (*optimization variable*). Hàm số $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ được gọi là *hàm mục tiêu* (*objective function*, các hàm mục tiêu trong Machine Learning thường được gọi là *hàm mất mát*). Các hàm số $f_i, h_j : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, 2, \dots, m; j = 1, 2, \dots, p$ được gọi là các *hàm ràng buộc* (hoặc đơn giản là *ràng buộc* - constraints). Tập hợp các điểm \mathbf{x} thỏa mãn các *ràng buộc* được gọi là *feasible set*. Mỗi điểm trong *feasible set* được gọi là *feasible point*, các điểm không trong *feasible set* được gọi là *infeasible points*.

Chú ý:

- Nếu bài toán là tìm giá trị lớn nhất thay vì nhỏ nhất, ta chỉ cần đổi dấu của $f_0(\mathbf{x})$.
- Nếu ràng buộc là *lớn hơn hoặc bằng*, tức $f_i(\mathbf{x}) \geq b_i$, ta chỉ cần đổi dấu của ràng buộc là sẽ có điều kiện *nhỏ hơn hoặc bằng* $-f_i(\mathbf{x}) \leq -b_i$.
- Các ràng buộc cũng có thể là *lớn hơn* hoặc *nhỏ hơn*.
- Nếu ràng buộc là *bằng nhau*, tức $h_j(\mathbf{x}) = 0$, ta có thể viết nó dưới dạng hai bất đẳng thức $h_j(\mathbf{x}) \leq 0$ và $-h_j(\mathbf{x}) \leq 0$. Trong một vài tài liệu, người ta bỏ các phương trình ràng buộc $h_j(\mathbf{x}) = 0$ đi.
- Trong bài viết này, \mathbf{x}, \mathbf{y} được dùng chủ yếu để ký hiệu các biến số, không phải là dữ liệu như trong các bài trước. Biến tối ưu chính là biến được ghi dưới dấu arg min. Khi viết một bài toán Tối Ưu, ta cần chỉ rõ biến nào cần được tối ưu, biến nào là cố định.

Các bài toán tối ưu, nhìn chung không có cách giải tổng quát, thậm chí có những bài chưa có lời giải. Hầu hết các phương pháp tìm nghiệm không chứng minh được nghiệm tìm được có phải là *global optimal* hay không, tức đúng là điểm làm cho hàm số đạt giá trị nhỏ nhất hay lớn nhất hay không. Thay vào đó, nghiệm thường là các *local optimal*, tức các *điểm cực trị*.

Để bắt đầu học Tối Ưu, chúng ta cần học một mảng rất quan trọng trong đó, có tên là *Tối Ưu Lồi* (convex optimization), trong đó *hàm mục tiêu* là một *hàm lồi* (convex function), *feasible set* là một *tập lồi* (convex set). Những tính chất đặc biệt về *local optimal* và *global optimal* của một *hàm lồi* khiến Tối Ưu Lồi trở nên cực kỳ quan trọng. Trong bài viết này, tôi sẽ giới thiệu tới các bạn các định nghĩa và tính chất cơ bản của *tập lồi* và *hàm lồi*. *Bài toán tối ưu lồi* (convex optimization problems) sẽ được đề cập trong bài tiếp theo.



Example of convex sets

Hình 16.1: Các ví dụ về convex sets.

16.2 Convex sets

16.2.1 Định nghĩa

Khái niệm về *convex sets* có lẽ không xa lạ với các bạn học sinh Việt Nam khi chúng ta đã nghe về *đa giác lồi*. *Lồi*, hiểu đơn giản là *phình ra ngoài*, hoặc *nhô ra ngoài*. Trong toán học, *bằng phẳng* cũng được coi là *lồi*.

Định nghĩa 1: Một tập hợp được gọi là *tập lồi* (convex set) nếu đoạn thẳng nối hai điểm bất kỳ trong tập hợp hợp đó nằm trọn vẹn trong tập hợp đó.

Một vài ví dụ về convex sets được cho trong Hình 16.1.

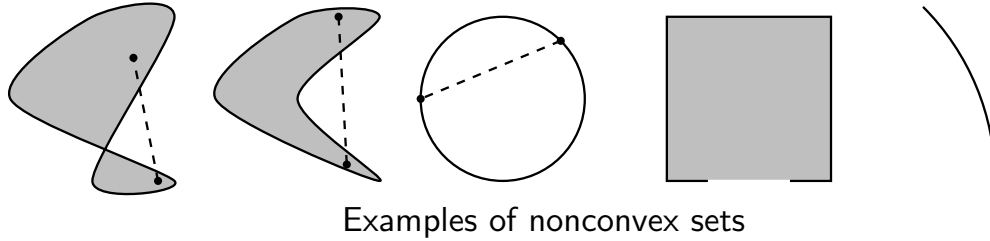
Các hình với đường biên màu đen thể hiện việc bao gồm cả biên, biên màu trắng thể hiện việc biên đó không nằm trong tập hợp đang xét. Đường hoặc đoạn thẳng cũng là một tập lồi theo định nghĩa phía trên.

Một vài ví dụ thực tế:

- Giả sử có một căn phòng có dạng hình *lồi*, nếu ta đặt một bóng đèn đủ sáng ở bất kỳ vị trí nào trong phòng, mọi điểm trong căn phòng đều được chiếu sáng.
- Nếu một đất nước có bản đồ dạng một hình *lồi* thì đường bay nối giữa hai thành phố bất kỳ trong đất nước đó đều nằm trọn vẹn trong không phận của nước đó. (Không như Việt Nam, muốn bay thẳng Hà Nội - Hồ Chí Minh phải bay qua không phận Campuchia).

Hình 16.2 minh họa một vài ví dụ về *nonconvex sets*, tức tập hợp mà không phải là lồi:

Ba hình đầu tiên không phải là lồi vì các đường nét đứt chứa nhiều điểm không nằm trong các tập đó. Hình thứ tư, hình vuông không có biên ở đáy, không phải là *tập lồi* vì đoạn thẳng nối hai điểm ở đáy có thể chứa phần ở giữa không thuộc tập đang xét (Nếu không có biên thì hình vuông vẫn là một *tập lồi*, nhưng biên nửa vời như ví dụ này thì hãy chú ý). Một đường cong bất kỳ cũng không phải là *tập lồi* vì dễ thấy đường thẳng nối hai điểm bất kỳ không thuộc đường cong đó.



Hình 16.2: Các ví dụ về nonconvex sets.

Để mô tả một *tập lồi* dưới dạng toán học, ta sử dụng:

Định nghĩa 2: Một tập hợp \mathcal{C} được gọi là *convex* nếu với hai điểm bất kỳ $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$, điểm $\mathbf{x}_\theta = \theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2$ cũng nằm trong \mathcal{C} với bất kỳ $0 \leq \theta \leq 1$.

Có thể thấy rằng, tập hợp các điểm có dạng $(\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2)$ chính là *đoạn thẳng* nối hai điểm \mathbf{x}_1 và \mathbf{x}_2 .

Với các định nghĩa này thì *toàn bộ không gian* là một *tập lồi* vì đoạn thẳng nào cũng nằm trong không gian đó. Tập rỗng cũng có thể coi là một trường hợp đặc biệt của *tập lồi*.

Dưới đây là một vài ví dụ hay gặp về *tập lồi*.

16.2.2 Ví dụ

Hyperplanes và halfspaces

Một **hyperplane** (siêu mặt phẳng) trong không gian n chiều là tập hợp các điểm thỏa mãn phương trình:

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \mathbf{a}^T \mathbf{x} = b$$

với $b, a_i, i = 1, 2, \dots, n$ là các số thực.

Hyperplanes là các *tập lồi*. Điều này có thể dễ dàng suy ra từ Định nghĩa 1. Với Định nghĩa 2, chúng ta cũng dễ dàng nhận thấy. Nếu:

$$\mathbf{a}^T \mathbf{x}_1 = \mathbf{a}^T \mathbf{x}_2 = b$$

thì với $0 \leq \theta \leq 1$ bất kỳ:

$$\mathbf{a}^T \mathbf{x}_\theta = \mathbf{a}^T (\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2) = \theta b + (1 - \theta) b = b$$

Một **halfspace** (nửa không gian) trong không gian n chiều là tập hợp các điểm thỏa mãn phương trình:

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = \mathbf{a}^T \mathbf{x} \leq b$$

với $b, a_i, i = 1, 2, \dots, n$ là các số thực.

Các halfspace cũng là các tập lồi, bạn đọc có thể dễ dàng nhận thấy theo Định nghĩa 1 hoặc chứng minh theo Định nghĩa 2.

Norm balls

Euclidean balls (hình tròn trong mặt phẳng, hình cầu trong không gian ba chiều) là tập hợp các điểm có dạng:

$$B(\mathbf{x}_c, r) = \{ \mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r = \mathbf{x}_c + r\mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1 \}$$

Theo Định nghĩa 1, chúng ta có thể *thấy* Euclidean balls là các tập lồi, nếu phải chứng minh, ta dùng Định nghĩa 2 và **các tính chất của norms**. Với $\mathbf{x}_1, \mathbf{x}_2$ bất kỳ thuộc $B(\mathbf{x}_c, r)$ và $0 \leq \theta \leq 1$ bất kỳ:

$$\begin{aligned} \|\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 - \mathbf{x}_c\|_2 &= \|\theta(\mathbf{x}_1 - \mathbf{x}_c) + (1-\theta)(\mathbf{x}_2 - \mathbf{x}_c)\|_2 \\ &\leq \theta\|\mathbf{x}_1 - \mathbf{x}_c\|_2 + (1-\theta)\|\mathbf{x}_2 - \mathbf{x}_c\|_2 \\ &\leq \theta r + (1-\theta)r = r \end{aligned}$$

Vậy nên $\mathbf{x}_\theta \in B(\mathbf{x}_c, r)$.

Euclidean ball sử dụng norm 2 làm khoảng cách. Nếu sử dụng norm bất kỳ là khoảng cách, ta vẫn được một *tập lồi*.

Khi sử dụng norm p:

$$\|\mathbf{x}\|_p = (\|x_1\|^p + \|x_2\|^p + \dots \|x_n\|^p)^{\frac{1}{p}} \quad (1)$$

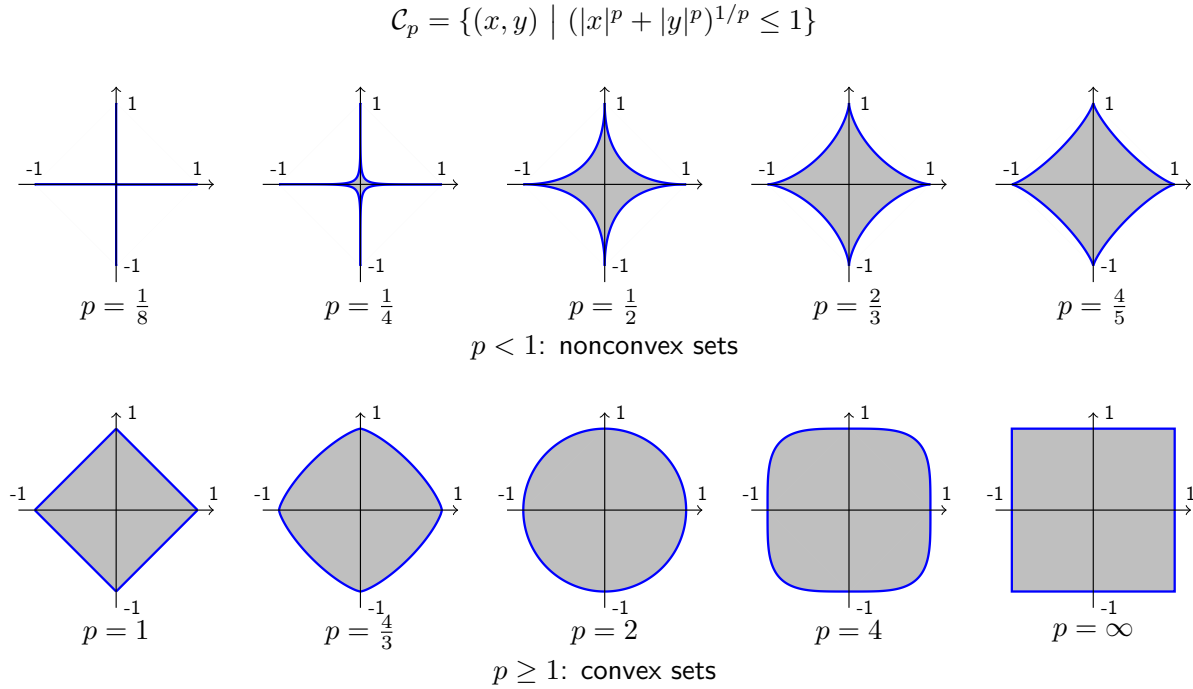
với p là một số thực bất kỳ không nhỏ hơn 1 ta cũng thu được các *tập lồi*.

Hình 16.3 minh họa tập hợp các điểm có tọa độ (x, y) trong không gian hai chiều thỏa mãn:

$$(|x|^p + |y|^p)^{1/p} \leq 1 \quad (1)$$

với hàng trên là các tập với $0 < p < 1$ (không phải norm) và hàng dưới tương ứng với $p \geq 1$.

Chúng ta có thể thấy rằng khi p nhỏ gần bằng 0, tập hợp các điểm thỏa mãn bất đẳng thức (1) gần như nằm trên các trục tọa độ và bị chặn trong đoạn $[0, 1]$. Quan sát này sẽ giúp ích



Hình 16.3: Hình dạng của các tập hợp bị chặn bởi pseudo-norms (hàng trên) và norm (hàng dưới).

cho các bạn khi làm việc với (giả) norm 0 sau này. Khi $p \rightarrow \infty$, các tập hợp hội tụ về hình vuông.

Đây cũng là một trong các lý do vì sao cần có điều kiện $p \geq 1$ khi định nghĩa norm.

Ellipsoids

Các ellipsoids (ellipse trong không gian nhiều chiều) cũng là các *tập lồi*. Thực chất, ellipsoids có mối quan hệ mật thiết tới [Khoảng cách Mahalanobis](#). Khoảng cách này vốn dĩ là một norm nên ta có thể chứng minh theo Định nghĩa 2 được tính chất lồi của các ellipsoids.

Mahalanobis norm của một vector $\mathbf{x} \in \mathbb{R}^n$ được định nghĩa là:

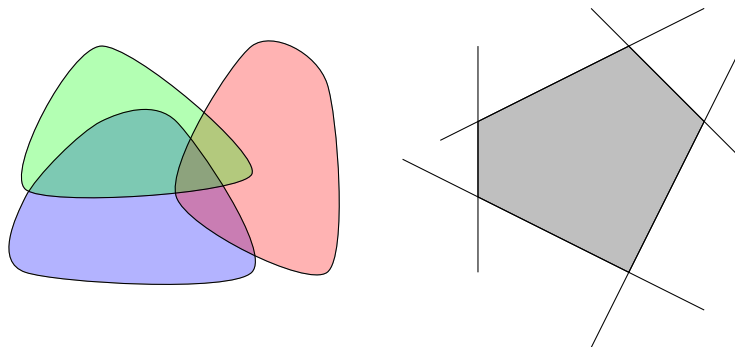
$$\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x}}$$

Với \mathbf{A} là một ma trận thỏa mãn:

$$\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n \quad (2)$$

Khi một ma trận \mathbf{A} thỏa mãn điều kiện (2), ta nói ma trận đó *xác định dương* (*positive definite*). Một ma trận là *xác định dương* nếu các *trị riêng* (eigenvalues) của nó là dương.

Nhân tiện, một ma trận \mathbf{B} được gọi là **nửa xác định dương** (*positive semidefinite*) nếu các *trị riêng* của nó là không âm. Khi đó $\mathbf{x}^T \mathbf{B} \mathbf{x} \geq 0, \forall \mathbf{x}$. Nếu dấu bằng xảy ra khi và chỉ khi



Hình 16.4: Trái: Giao của các tập lồi là một tập lồi. Phải: giao của các hyperplanes và hafspaces là một tập lồi và được gọi là polyhedron (số nhiều là polyhedra).

$\mathbf{x} = 0$ thì ta nói ma trận đó *xác định dương*. Trong biểu thức (2), vì ma trận \mathbf{A} có nghịch đảo nên mọi *trị riêng* của nó phải khác không. Vì vậy, \mathbf{A} là một ma trận *xác định dương*.

Một ma trận \mathbf{A} là *xác định dương* hoặc *nửa xác định dương* sẽ được ký hiệu lần lượt như sau:

$$\mathbf{A} \succ 0, \quad \mathbf{A} \succeq 0.$$

Cũng lại nhân tiện, khoảng cách Mahalanobis có liên quan đến *khoảng cách từ một điểm tới một phân phối xác suất* (from a point to a distribution).

16.2.3 Giao của các tập lồi là một tập lồi.

Việc này có thể nhận dễ nhận thấy với Hình 16.4 (trái). Giao của hai trong ba hoặc cả ba tập lồi đều là các tập lồi.

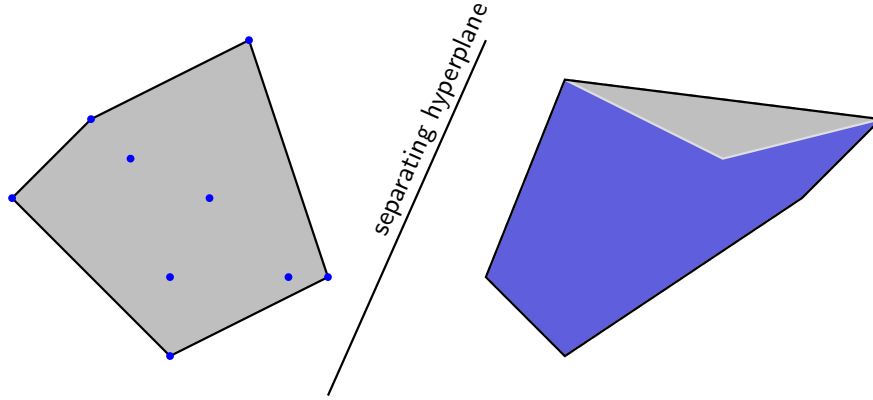
Việc chứng minh việc này theo Định nghĩa 2 cũng không khó. Nếu $\mathbf{x}_1, \mathbf{x}_2$ thuộc vào giao của các tập lồi, tức thuộc tất cả các tập lồi đã cho, thì $\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$ cũng thuộc vào tất cả các tập lồi, tức thuộc vào giao của chúng!

Từ đó suy ra giao của các *hafspaces* và các *hyperplanes* cũng là một tập lồi. Trong không gian hai chiều, tập lồi này chính là *đa giác lồi*, trong không gian ba chiều, nó có tên là *đa diện lồi*.

Trong không gian nhiều chiều, giao của các *hafspaces* và *hyperplanes* được gọi là **polyhedra**.

Giả sử có m *hafspaces* và p *hyperplanes*. Mỗi một *hafspace*, theo như đã trình bày phía trên, có thể viết dưới dạng $\mathbf{a}_i^T \mathbf{x} \leq b_i, \forall i = 1, 2, \dots, m$. Mỗi một *hyperplane* có thể viết dưới dạng: $\mathbf{c}_i^T \mathbf{x} = d_i, \forall i = 1, 2, \dots, p$.

Vậy nếu đặt $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]$, $\mathbf{b} = [b_1, b_2, \dots, b_m]^T$, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p]$ và $\mathbf{d} = [d_1, d_2, \dots, d_p]^T$, ta có thể viết polyhedra dưới dạng tập hợp các điểm \mathbf{x} thỏa mãn:



Hình 16.5: Trái: Giao của các tập lồi là một tập lồi. Phải: giao của các hyperplanes và halfspaces là một tập lồi và được gọi là polyhedron (số nhiều là polyhedra).

$$\mathbf{A}^T \mathbf{x} \preceq \mathbf{b}, \quad \mathbf{C}^T \mathbf{x} = \mathbf{d}$$

trong đó \preceq là *element-wise*, tức mỗi phần tử trong vế trái nhỏ hơn hoặc bằng phần tử tương ứng trong vế phải.

16.2.4 Convex combination và Convex hulls

Một điểm được gọi là **convex combination** (*tổ hợp lồi*) của các điểm $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ nếu nó có thể viết dưới dạng:

$$\mathbf{x} = \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + \dots + \theta_k \mathbf{x}_k, \quad \text{with } \theta_1 + \theta_2 + \dots + \theta_k = 1$$

Convex hull của một **tập hợp bất kỳ** là tập hợp tất cả các điểm là *convex combination* của tập hợp đó. *Convex hull* là một *convex set*. *Convex hull* của một *convex set* là chính nó. Một cách dễ nhớ, *convex hull* của một tập hợp là một *convex set* **nhỏ nhất** chứa tập hợp đó. Khái niệm **nhỏ nhất** rất khó định nghĩa, nhưng nó cũng là một cách nhớ trực quan.

Hai tập hợp được gọi là *linearly separable* nếu các *convex hulls* của chúng không có điểm chung.

Trong Hình 16.5, convex hull của các điểm màu xanh là vùng màu xám bao với các đa giác lồi. Ở Hình 16.5 phải, vùng màu xám nằm dưới vùng màu xanh.

Định lý siêu phẳng phân chia (Separating hyperplane theorem): Định lý này nói rằng nếu hai *tập lồi không rỗng* \mathcal{C}, \mathcal{D} là *disjoint* (không giao nhau), thì tồn tại vector \mathbf{a} và số b sao cho:

$$\mathbf{a}^T \mathbf{x} \leq b, \forall \mathbf{x} \in \mathcal{C}, \quad \mathbf{a}^T \mathbf{x} \geq b, \forall \mathbf{x} \in \mathcal{D}$$

Tập hợp tất cả các điểm \mathbf{x} thỏa mãn $\mathbf{a}^T \mathbf{x} = b$ chính là một hyperplane. Hyperplane này được gọi là *separating hyperplane*.

Ngoài ra còn nhiều tính chất thú vị của các tập lồi và các phép toán bảo toàn chính chất *lồi* của một tập hợp, các bạn được khuyến khích đọc thêm Chương 2 của cuốn Convex Optimization trong phần tài liệu tham khảo.

16.3 Convex functions

Hẳn các bạn đã nghe tới khái niệm này khi ôn thi đại học môn toán. Khái niệm hàm lồi có quan hệ tới đạo hàm bậc hai và **Bất đẳng thức Jensen** (nếu bạn chưa nghe tới phần này, không sao, bây giờ bạn sẽ biết).

16.3.1 Định nghĩa

Để trực quan, trước hết ta xem xét các hàm 1 biến, đồ thị của nó là một đường trong một mặt phẳng. Một hàm số được gọi là *lồi* nếu **tập xác định của nó là một tập lồi** và nếu ta nối hai điểm bất kỳ trên đồ thị hàm số đó, ta được một đoạn thẳng nằm về phía trên hoặc nằm trên đồ thị (xem Hình 16.6).

Tập xác định (domain) của một hàm số $f(\cdot)$ thường được ký hiệu là $\text{dom} f$.

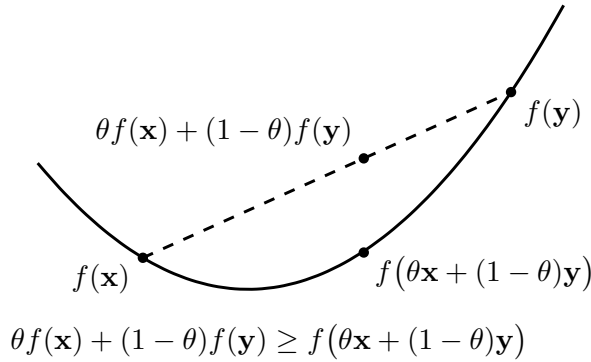
Định nghĩa theo toán học:

Định nghĩa convex function: Một hàm số $f : \mathbb{R}^n \rightarrow \mathbb{R}$ được gọi là một *hàm lồi* (convex function) nếu $\text{dom} f$ là một *tập lồi*, và:

$$f(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y})$$

với mọi $\mathbf{x}, \mathbf{y} \in \text{dom} f, 0 \leq \theta \leq 1$.

Điều kiện $\text{dom} f$ là một *tập lồi* là rất quan trọng, vì nếu không có nó, ta không định nghĩa được $f(\theta \mathbf{x} + (1 - \theta) \mathbf{y})$.



Hình 16.6: Định nghĩa hàm lồi. Diễn đạt bằng lời, một hàm số là lồi nếu đoạn thẳng nối 2 điểm bất kỳ trên đồ thị của nó *không nằm dưới* đồ thị đó.

Một hàm số f được gọi là **concave** (nếu bạn muốn dịch là *lõm* cũng được, tôi không thích cách dịch này) nếu $-f$ là **convex**. Một hàm số có thể không thuộc hai loại trên. Các hàm tuyến tính vừa *convex*, vừa *concave*.

Định nghĩa strictly convex function: (tiếng Việt có một số tài liệu gọi là *hàm lồi mạnh* hoặc *hàm lồi chặt*) Một hàm số $f : \mathbb{R}^n \rightarrow \mathbb{R}$ được gọi là *strictly convex* nếu $\text{dom} f$ là một tập lồi, và:

$$f(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) < \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y})$$

với mọi $\mathbf{x}, \mathbf{y} \in \text{dom} f, \mathbf{x} \neq \mathbf{y}, 0 < \theta < 1$.

Tương tự với định nghĩa **strictly concave**.

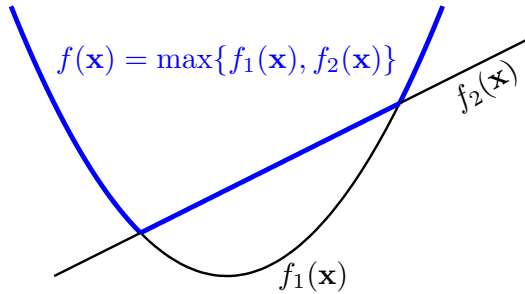
Đây là một điểm quan trọng: Nếu một hàm số là *strictly convex* và có điểm cực trị, thì điểm cực trị đó là duy nhất và cũng là *global minimum*.

16.3.2 Các tính chất cơ bản

- Nếu $f(\mathbf{x})$ là *convex* thì $af(\mathbf{x})$ là *convex* nếu $a > 0$ và là *concave* nếu $a < 0$. Điều này có thể suy ra trực tiếp từ định nghĩa.
- Tổng của hai *hàm lồi* là một *hàm lồi*, với tập xác định là giao của hai tập xác định kia (nhắc lại rằng giao của hai tập lồi là một tập lồi)
- **Pointwise maximum and supremum:** Nếu các hàm số f_1, f_2, \dots, f_m là *convex* thì:

$$f(\mathbf{x}) = \max f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})$$

cũng là *convex* trên tập xác định là giao của tất cả các tập xác định của các hàm số trên. Hàm \max phía trên cũng có thể thay thế bằng *hàm sup*. Tính chất này có thể chứng minh được theo Định nghĩa. Bạn cũng có thể nhận ra dựa vào hình ví dụ dưới đây. Mọi đoạn thẳng nối hai điểm bất kỳ trên đường màu xanh đều *không nằm dưới* đường màu xanh.



Hình 16.7: Ví dụ về Pointwise maximum. Maximum của các hàm lồi là một hàm lồi.

16.3.3 Ví dụ

Các hàm một biến

Các ví dụ về các *convex functions* một biến:

- Hàm $y = ax + b$ là một *hàm lồi* vì đường nối hai điểm bất kỳ nằm trên chính đồ thị đó.
- Hàm $y = e^{ax}$ với $a \in \mathbb{R}$ bất kỳ.
- Hàm $y = x^a$ trên tập các số thực dương và $a \geq 1$ hoặc $a \leq 0$.
- Hàm *negative entropy* $y = x \log x$ trên tập các số thực dương.

Dưới đây là đồ thị của một vài *convex functions*:

Các ví dụ về các *concave functions* một biến:

- Hàm $y = ax + b$ là một *concave function* vì $-y$ là một *convex function*.
- Hàm $y = x^a$ trên tập số dương và $0 \leq a \leq 1$.
- Hàm logarithm $y = \log(x)$ trên tập các số dương.

Dưới đây là đồ thị của một vài *concave functions*:

Affine functions

Các hàm số dạng $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + b$ vừa là convex, vừa là concave.

Khi biến là một ma trận \mathbf{X} , các hàm affine được định nghĩa có dạng:

$$f(\mathbf{X}) = \text{trace}(\mathbf{A}^T \mathbf{X}) + b$$

trong đó trace là hàm số tính tổng các giá trị trên đường chéo của một ma trận vuông, \mathbf{A} là một ma trận có cùng chiều với \mathbf{X} (để đảm bảo phép nhân ma trận thực hiện được và kết quả là một ma trận vuông).

Quadratic forms

Hàm bậc hai một biến có dạng $f(x) = ax^2 + bx + c$ là convex nếu $a > 0$, là concave nếu $a < 0$.

Với biến là một vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$, một quadratic form là một hàm số có dạng:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$

Với \mathbf{A} thường là một ma trận đối xứng, tức $a_{ij} = a_{ji}, \forall i, j$, có số hàng bằng số phần tử của \mathbf{x} , \mathbf{b} là một ma trận bất kỳ cùng chiều với \mathbf{x} và c là một hằng số bất kỳ.

Nếu \mathbf{A} là một ma trận (nửa) xác định dương thì $f(\mathbf{x})$ là một *convex function*.

Nếu \mathbf{A} là một ma trận (nửa) xác định âm, tức $\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0, \forall \mathbf{x}$, thì $f(\mathbf{x})$ là một *concave function*.

Các bạn có thể tìm đọc về ma trận xác định dương và các tính chất của nó trong sách Đại số tuyến tính bất kỳ. Nếu bạn gặp nhiều khó khăn trong phần này, hãy đọc lại kiến thức về Đại số tuyến tính, rất rất quan trọng trong Tối Ưu và Machine Learning.

Hàm mất mát trong Linear Regression có dạng:

$$\begin{aligned} \mathcal{L}(\mathbf{w}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 = \frac{1}{2} (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}) \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w} - \mathbf{y}^T \mathbf{X} \mathbf{w} + \frac{1}{2} \mathbf{y}^T \mathbf{y} \end{aligned}$$

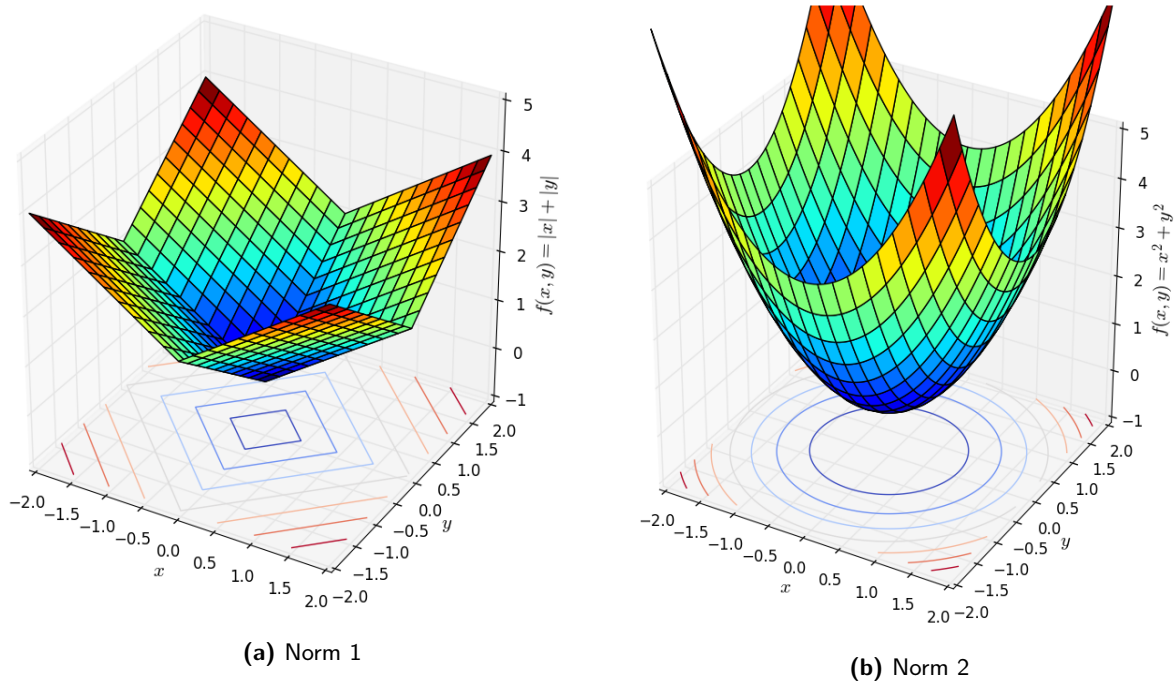
vì $\mathbf{X}^T \mathbf{X}$ là một ma trận xác định dương, hàm mất mát của Linear Regression chính là một convex function.

Norms

Vâng, lại là norms. Một hàm số bất kỳ thỏa mãn [ba điều kiện của norm](#) đều là một *convex function*. Bạn đọc có thể chứng minh điều này bằng định nghĩa.

Hình 16.8 minh họa hai ví dụ về norm 1 (trái) và norm 2 (phải) với số chiều là 2 (chiều thứ ba trong hình dưới đây là giá trị của hàm số).

Nhận thấy rằng các bề mặt này đều có *một đáy duy nhất* tương ứng với gốc tọa độ (đây chính là điều kiện đầu tiên của norm). Các hàm *strictly convex* khác cũng có dạng tương tự,



Hình 16.8: Ví dụ về mặt của các norm hai biến.

tức có một *đáy* duy nhất. Điều này cho thấy nếu ta *thả một hòn bi* ở vị trí bất kỳ trên các bề mặt này, cuối cùng nó sẽ *lăn* về đáy. Nếu liên tưởng tới thuật toán [Gradient Descent](#) thì việc áp dụng thuật toán này vào các bài toán không ràng buộc với *hàm mục tiêu* là *strictly convex* (và giả sử là khả vi, tức có đạo hàm) sẽ cho kết quả rất tốt nếu *learning rate* không quá lớn. Đây chính là một trong các lý do vì sao các *convex functions* là quan trọng, cũng là lý do vì sao tôi dành bài viết này chỉ để nói về *convexity*. (Bạn đọc được khuyến khích đọc hai bài về [Gradient Descent](#) trong blog này).

Tiện đây, tôi cũng lấy thêm hai ví dụ về các hàm không phải convex (cũng không phải concave). Hàm thứ nhất $f(x, y) = x^2 - y^2$ là một hyperbolic, hàm thứ hai $f(x, y) = \frac{1}{10}(x^2 + 2y^2 - 2\sin(xy))$.

Contours - level sets Với các hàm số phức tạp hơn, khi vẽ các mặt trong không gian ba chiều sẽ khó tưởng tượng hơn, tức khó nhìn được tính *convexity* của nó. Một phương pháp thường được sử dụng là dùng *contours* hay *level sets*. Tôi cũng đã đề cập đến khái niệm này trong Bài Gradient Descent, phần [đường đồng mức](#).

Contours là cách mô tả các mặt trong không gian ba chiều bằng cách chiếu nó xuống không gian hai chiều. Trong không gian hai chiều, các điểm thuộc cùng một *đường* tương ứng với các điểm làm cho hàm số có giá trị bằng nhau. Mỗi *đường* đó còn được gọi là một *level set*. Trong Hình 8 và Hình 9, các đường của các mặt lên mặt phẳng Oxy chính là các *level sets*. Một cách hiểu khác, mỗi đường *level set* là một *vết cắt* nếu ta cắt các bề mặt bởi một mặt phẳng song song với mặt phẳng Oxy .

Khi thể hiện một hàm số hai biến để kiểm tra tính convexity của nó, hoặc để tìm điểm cực trị của nó, người ta thường vẽ *contours* thay vì vẽ các mặt trong không gian ba chiều. Dưới đây là một vài ví dụ về contours:

Các đường màu càng xanh đậm thì tương ứng với các giá trị càng nhỏ, các đường màu càng đỏ đậm thì tương ứng các giá trị càng lớn.

Ở hàng trên, các đường *level sets* là các đường khép kín (closed). Khi các đường kín này tập trung nhỏ dần ở một điểm thì các điểm đó là các điểm cực trị. Với các *convex functions* như trong ba ví dụ này, chỉ có 1 điểm cực trị và đó cũng là điểm làm cho hàm số đạt giá trị nhỏ nhất (global optimal). Nếu để ý, bạn sẽ thấy các đường khép kín này tạo thành một *vùng lồi*!

Ở hàng dưới, các đường không phải khép kín. Hình bên trái tương ứng với một hàm tuyến tính $f(x, y) = x + y$ và đó là một *convex function*. Hình ở giữa cũng là một *convex function* (bạn có thể chứng minh điều này sau khi tính đạo hàm bậc hai, tôi sẽ nói ở phía dưới) nhưng các level sets là các *đường không kín*. Hàm này có log nên tập xác định là góc phần tư thứ nhất tương ứng với các tọa độ dương (chú ý rằng tập hợp các điểm có tọa độ dương cũng là một *tập lồi*). Các *đường không kín* này nếu kết hợp với trục Ox, Oy sẽ tạo thành biên của các *tập lồi*. Hình cuối cùng là contours của một hàm hyperbolic, hàm này không phải là *hàm lồi*.

16.3.4 α — sublevel sets

Định nghĩa: α —sublevel set của một hàm số $f : \mathbb{R}^n \rightarrow \mathbb{R}$ được định nghĩa là:

$$\mathcal{C}_\alpha = \{\mathbf{x} \in \text{dom} f \mid f(\mathbf{x}) \leq \alpha\}$$

Tức tập hợp các điểm trong tập xác định của f mà tại đó, f đạt giá trị nhỏ hơn hoặc bằng α .

Quay lại với Hình 11, hàng trên, các α —sublevel sets chính là phần bị bao bởi các level sets.

Ở hàng dưới, bên trái, các α —sublevel sets chính là phần nửa mặt phẳng phía dưới xác định bởi các đường thẳng level sets. Ở hình giữa, các α —sublevel sets chính là các vùng bị giới hạn bởi các trục tọa độ và các level sets.

Hàng dưới, bên phải, các α —sublevel sets hơi khó tưởng tượng chút. Với $\alpha > 0$, các level sets là các đường màu vàng hoặc đỏ. Các α —sublevel sets tương ứng là phần *bị bóp vào trong*, giới hạn bởi các đường đỏ cùng màu. Các vùng này, có thể dễ nhận thấy, là *không lồi*.

<hr> **Định lý:** Nếu một hàm số là lồi thì mọi α -sublevel sets của nó là lồi. Ngược lại chưa chắc đã đúng, tức nếu các α -sublevel sets của một hàm số là lồi thì hàm số đó chưa chắc đã lồi. <hr>

Điều này chỉ ra rằng nếu tồn tại một giá trị α sao cho một α -sublevel set của một hàm số là không lồi, thì hàm số đó là không lồi (không lồi nhưng không có nghĩa là concave, chú ý). Vậy nên Hyperbolic không phải là hàm lồi.

Các ví dụ ở hình 11, trừ hình cuối cùng, đều tương ứng với các hàm lồi.

Một ví dụ về việc một hàm số không convex nhưng mọi α -sublevel sets là convex là hàm $f(x, y) = -e^{x+y}$. Hàm này có mọi α -sublevel sets là nửa mặt phẳng - là convex, nhưng nó không phải là convex (trong trường hợp này nó là concave).

Dưới đây là một ví dụ khác về việc một hàm số có mọi α -sublevel sets là lồi nhưng không phải hàm lồi.

Mọi α -sublevel sets của hàm số này đều là các hình tròn - convex nhưng hàm số đó không phải là lồi. Vì có thể tìm được hai điểm trên mặt này sao cho đoạn thẳng nối hai điểm nằm hoàn toàn phía dưới của mặt (một điểm ở cánh và 1 điểm ở đáy chẳng hạn).

Những hàm số có tập xác định là một tập lồi và có mọi α -sublevel sets là lồi được gọi chung là *quasiconvex*. Mọi convex function đều là quasiconvex nhưng ngược lại không đúng. Định nghĩa chính thức của *quasiconvex function* được phát biểu như sau: <hr> **Quasiconvex function:** Một hàm số $f : \mathcal{C} \rightarrow \mathbb{R}$ với \mathcal{C} là một tập con lồi của \mathbb{R}^n được gọi là *quasiconvex* nếu với mọi $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ và mọi $\theta \in [0, 1]$, ta có:

$$f(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$$

Định nghĩa này khác với định nghĩa về *convex function* một chút.

16.3.5 Kiểm tra tính chất lồi dựa vào đạo hàm.

Có một cách để nhận biết một hàm số khả vi có là hàm lồi hay không dựa vào các đạo hàm bậc nhất hoặc đạo hàm bậc hai của nó.

First-order condition

Trước hết chúng ta định nghĩa phương trình đường (mặt) tiếp tuyến của một hàm số f khả vi tại một điểm nằm trên đồ thị (mặt) của hàm số đó $(\mathbf{x}_0, f(\mathbf{x}_0))$. Với hàm một biến, bạn đọc đã quen thuộc:

$$y = f'(x_0)(x - x_0) + f(x_0)$$

Với hàm nhiều biến, đặt $\nabla f(\mathbf{x}_0)$ là gradient của hàm số f tại điểm \mathbf{x}_0 , phương trình mặt tiếp tuyến được cho bởi:

$$y = \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + f(\mathbf{x}_0)$$

<hr> **First-order condition** nói rằng: Giả sử hàm số f có tập xác định là một tập lồi, có đạo hàm tại mọi điểm trên tập xác định đó. Khi đó, hàm số f là *lồi* **nếu và chỉ nếu** với mọi \mathbf{x}, \mathbf{x}_0 trên tập xác định của hàm số đó, ta có:

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) \quad (6)$$

<hr>

Tương tự như thế, một hàm số là *strictly convex* nếu dấu bằng trong (6) xảy ra khi và chỉ khi $\mathbf{x} = \mathbf{x}_0$.

Nói một cách trực quan hơn, một hàm số là lồi nếu đường (mặt) tiếp tuyến tại một điểm bất kỳ trên đồ thị (mặt) của hàm số đó **nằm dưới** đồ thị (mặt) đó. (Đừng quên điều kiện về tập xác định là lồi) Dưới đây là ví dụ về *hàm lồi* và *hàm không lồi*. Hàm bên trái là một hàm lồi. Hàm bên phải không phải là hàm lồi vì đồ thị của nó vừa nằm trên, vừa nằm dưới tiếp tuyến.

(*iff* là viết tắt của *if and only if*)

Ví dụ: Nếu ma trận đối xứng \mathbf{A} là *xác định dương* thì hàm số $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ là *hàm lồi*.

Chứng minh: Đạo hàm bậc nhất của hàm số trên là:

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x}$$

Vậy *first-order condition* có thể viết dưới dạng (chú ý rằng \mathbf{A} là một ma trận đối xứng):

$$\begin{aligned} \mathbf{x}^T \mathbf{A} \mathbf{x} &\geq 2(\mathbf{A}\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0^T \mathbf{A} \mathbf{x}_0 \\ \Leftrightarrow \mathbf{x}^T \mathbf{A} \mathbf{x} &\geq 2\mathbf{x}_0^T \mathbf{A} \mathbf{x} - \mathbf{x}_0^T \mathbf{A} \mathbf{x}_0 \\ \Leftrightarrow (\mathbf{x} - \mathbf{x}_0)^T \mathbf{A} (\mathbf{x} - \mathbf{x}_0) &\geq 0 \end{aligned}$$

Bất đẳng thức cuối cùng là đúng dựa trên định nghĩa của một ma trận *xác định dương*. Vậy hàm số $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ là *hàm lồi*.

First-order condition ít được sử dụng để tìm tính chất lồi của một hàm số, thay vào đó, người ta thường dùng *Second-order condition* với các hàm có đạo hàm tới bậc hai.

Second-order condition

Với hàm nhiều biến, tức biến là một vector, giả sử có chiều là d , đạo hàm bậc nhất của nó là một vector cũng có chiều là d . Đạo hàm bậc hai của nó là một ma trận vuông có chiều

là $d \times d$. Đạo hàm bậc hai của hàm số $f(\mathbf{x})$ được ký hiệu là $\nabla^2 f(\mathbf{x})$. Đạo hàm bậc hai còn được gọi là *Hessian*.

Second-order condition: Một hàm số có đạo hàm bậc hai là *convex* nếu $\text{dom} f$ là *convex* và Hessian của nó là một ma trận *nửa xác định dương* với mọi \mathbf{x} trong tập xác định:

$$\nabla^2 f(\mathbf{x}) \succeq 0.$$

Nếu Hessian là một ma trận *xác định dương* thì hàm số đó *strictly convex*. Tương tự, nếu Hessian là một ma trận *xác định âm* thì hàm số đó là *strictly concave*.

Với hàm số một biến $f(x)$, điều kiện này tương đương với $f''(x) \geq 0$ với mọi x thuộc tập xác định (và tập xác định là *lồi*).

Ví dụ:

- Hàm *negative entropy* $f(x) = x \log(x)$ là *strictly convex* vì tập xác định là $x > 0$ là một tập *lồi* và $f''(x) = 1/x$ là một số dương với mọi x thuộc tập xác định.
- Hàm $f(x) = x^2 + 5 \sin(x)$ không là hàm *lồi* vì đạo hàm bậc hai $f''(x) = 2 - 5 \sin(x)$ có thể nhận giá trị âm.
- Hàm *cross entropy* là một hàm *strictly convex*. Xét ví dụ đơn giản với chỉ hai xác suất x và $1 - x$ với a là một hằng số thuộc đoạn $[0, 1]$ và $0 < x < 1$: $f(x) = -(a \log(x) + (1 - a) \log(1 - x))$ có đạo hàm bậc hai là $\frac{a}{x^2} + \frac{1-a}{(1-x)^2}$ là một số dương.
- Nếu \mathbf{A} là một ma trận xác định dương thì $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}$ là *lồi* vì Hessian của nó chính là \mathbf{A} là một ma trận xác định dương.
- Xét hàm số *negative entropy* với hai biến: $f(x, y) = x \log(x) + y \log(y)$ trên tập các giá trị dương của x và y . Hàm số này có đạo hàm bậc nhất là $[\log(x) + 1, \log(y) + 1]^T$ và Hessian là:

là một ma trận đường chéo với các thành phần trên đường chéo là dương nên là một ma trận xác định dương. Vậy *negative entropy* là một hàm *strictly convex*. (Chú ý rằng một ma trận là xác định dương nếu các trị riêng của nó đều dương. Với một ma trận là ma trận đường chéo thì các trị riêng của nó chính là các thành phần trên đường chéo.)

Ngoài ra còn nhiều tính chất thú vị của các hàm *lồi*, các bạn được khuyến khích đọc thêm Chương 3 của cuốn Convex Optimization trong phần tài liệu tham khảo.

16.4 Tóm tắt

- Machine Learning và Optimization có quan hệ mật thiết với nhau. Trong Optimization, Convex Optimization là quan trọng nhất. Một bài toán là convex optimization nếu *hàm mục tiêu* là convex và tập hợp các điểm thỏa mãn các điều kiện ràng buộc là một *convex set*.
- Trong *convex set*, mọi đoạn thẳng nối hai điểm bất kỳ trong tập đó sẽ nằm hoàn toàn trong tập đó. Tập hợp các giao điểm của các *convex sets* là một *convex set*.
- Một hàm số là *convex* nếu đoạn thẳng nối hai điểm bất kỳ trên đồ thị hàm số đó không nằm dưới đồ thị đó.
- Một hàm số khả vi là *convex* nếu tập xác định của nó là *convex* và đường (mặt) tiếp tuyến *không nằm phía trên* đồ thị (bề mặt) của hàm số đó.
- Các norms là các hàm lồi, được sử dụng nhiều trong tối ưu.

16.5 Tài liệu tham khảo

- [1] [Convex Optimization](#) – Boyd and Vandenberghe, Cambridge University Press, 2004.

Index

convex combination, [8](#)
convex hull, [8](#)

Separating hyperplane theorem, [9](#)