LECTURE 23: NON-LINEAR PROGRAMMING (II)

1. The True Second-Derivative Test

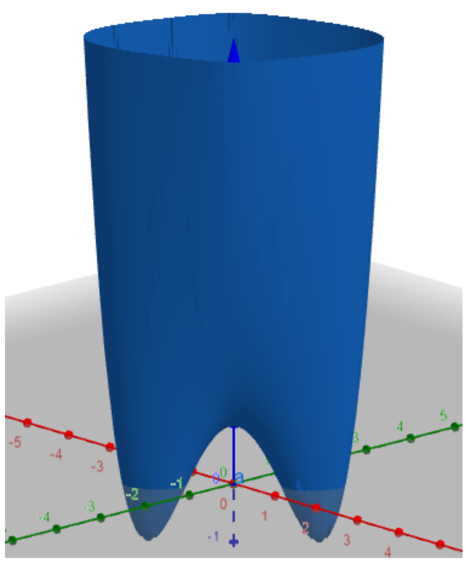
Video: The True Second-Derivative Test

How to find the local max/min of a function in higher dimensions.

Example 1:

Find the local max/min/saddle points of

$$f(x_1, x_2) = (x_1)^4 + (x_2)^4 - 4(x_1)(x_2) + 1$$



Date: Thursday, December 1, 2022.

STEP 1: Critical Points

Recall: (1D) c is a CP of f if $f'(c) = 0$ (or f' doesn't exist)

> **Definition: (Critical Point)**
>
> (a, b) is a **critical point** of f if $f_{x_1}(a, b) = 0$ and $f_{x_2}(a, b) = 0$

This is sometimes written as $\nabla f(a, b) = (0, 0)$

$$\begin{aligned} f_{x_1} =& 4 (x_1)^3 - 4x_2 = 0 \\ f_{x_2} =& 4 (x_2)^3 - 4x_1 = 0 \end{aligned}$$

The first equation gives $x_2 = (x_1)^3$ and plugging this into the second equation, we get

$$4 \left((x_1)^3 \right)^3 - 4x_1 = 0 \Rightarrow 4 (x_1)^9 - 4x_1 = 0 \Rightarrow 4x_1 \left((x_1)^8 - 1 \right) = 0$$

This gives us $x_1 = 0$ or $(x_1)^8 = 1$ that is, $x_1 = 0$ or $x_1 = 1$ or $x_1 = -1$

Case 1: $x_1 = 0$ then $x_2 = (x_1)^3 = 0^3 = 0$ which gives us $(0, 0)$

Case 2: $x_1 = 1$ then $x_2 = (x_2)^3 = 1^3 = 1$ which gives $(1, 1)$

Case 3: $x_1 = -1$ then $x_2 = (-1)^3 = -1$ which gives $(-1, -1)$

Conclusion: The critical points are $(0, 0), (1, 1), (-1, -1)$

STEP 2: Second derivatives

Recall: (1D) $f''(c) > 0 \Rightarrow$ local min, $f''(c) < 0 \Rightarrow$ local max

Here there are 4 second derivatives, so let's put them in a matrix

Definition:

$$D^2 f(x_1, x_2) = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} \\ f_{x_2x_1} & f_{x_2x_2} \end{bmatrix}$$

$$\begin{aligned} f_{x_1x_1} &= \left(4(x_1)^3 - 4x_2\right)_{x_1} = 12(x_1)^2 \\ f_{x_1x_2} &= \left(4(x_1)^3 - 4x_2\right)_{x_2} = -4 \\ f_{x_2x_1} &= \left(4(x_2)^3 - 4x_1\right)_{x_1} = -4 \\ f_{x_2x_2} &= \left(4(x_2)^3 - 4x_1\right)_{x_2} = 12(x_2)^2 \end{aligned}$$

$$D^2 f(x_1, x_2) = \begin{bmatrix} 12(x_1)^2 & -4 \\ -4 & 12(x_2)^2 \end{bmatrix}$$

This matrix is symmetric, this is because of Clairaut: $f_{x_1x_2} = f_{x_2x_1}$

Case 1: $(0, 0)$

$$D^2 f(0,0) = \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix} = A$$

The correct analog of $f''(c) > 0$ has to do with eigenvalues!

Note: For a review of eigenvalues, check out the following video:

Video: Eigenvalues

Eigenvalues:

$$\left|A - \lambda I\right| = \begin{vmatrix} -\lambda & -4 \\ -4 & -\lambda \end{vmatrix} = (-\lambda)^2 - (-4)^2 = \lambda^2 - 16 = 0 \Rightarrow \lambda = \pm 4$$

The eigenvalues have mixed signs, both positive and negative

The True Second Derivative Test:

Suppose (a, b) is a CP of f, then:

(1) If the eigenvalues of $D^2 f(a, b)$ are all positive, then f has a local min at (a, b)

(2) If the eigenvalues are all negative, then f has a local max at (a, b)

(3) If the eigenvalues have mixed sign, then f has a saddle point at (a, b)

And if one eigenvalue is 0, the test is inconclusive

Here f has a saddle at $(0, 0)$

Case 2: $(1, 1)$

$$D^2 f(1, 1) = \begin{bmatrix} 12 & -4 \\ -4 & 12 \end{bmatrix} = A$$

$$\left|A - \lambda I\right| = \begin{vmatrix} 12 - \lambda & -4 \\ -4 & 12 - \lambda \end{vmatrix} = (12 - \lambda)^2 - (-4)^2 = (\lambda - 12)^2 - 16 = 0$$

$$\lambda - 12 = \pm 4 \Rightarrow \lambda = 12 + 4 \text{ or } 12 - 4 \Rightarrow \lambda = 8 \text{ or } 16$$

All the eigenvalues are positive, so f has a local min at $(1, 1)$

Case 3: $(-1,-1)$

$$D^2 f(-1,-1) = \begin{bmatrix} 12 & -4 \\ -4 & 12 \end{bmatrix} \text{ same matrix}$$

f has a local min at $(-1,-1)$

Conclusion:

$$\begin{aligned} \text{Saddle Point: } & f(0,0) = 1 \\ \text{Local Min: } & f(1,1) = 1 + 1 - 4 + 1 = -1 \\ \text{Local Min: } & f(-1,-1) = 1 + 1 - 4(-1)(-1) + 1 = -1 \end{aligned}$$

Note: All of this works in higher dimensions as well!

2. Convexity

Question: How to show that f is convex?

Recall: (1D) f is convex (concave up) if $f''(x) > 0$ for all x

Example 2:

Is the following function convex?

$$f(x_1, x_2) = (x_1)^4 + (x_2)^4 - 4(x_1)(x_2) + 1$$

$$D^2 f(x_1, x_2) = \begin{bmatrix} 12(x_1)^2 & -4 \\ -4 & 12(x_2)^2 \end{bmatrix}$$

First Method:

Definition:

IF $D^2 f$ exists, then f is **convex** if for every (x_1, x_2), $D^2 f(x_1, x_2)$ has only positive eigenvalues.

We've seen that $D^2 f(0,0)$ has eigenvalues $\lambda = \pm 4$, so the answer is **NO**

That said, this method is not practical at all to check for convexity, because the eigenvalues here depend on x_1 and x_2

Faster Method: This is better explained in 3 dimensions

Example 3:

Is the following function convex?

$f(x_1, x_2, x_3) = (x_1)^2 + 2(x_2)^2 + 3(x_3)^2 + 2(x_1)(x_2) + 2(x_1)(x_3) + 3$

$$D^2 f(x_1, x_2, x_3) = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & f_{x_1x_3} \\ f_{x_2x_1} & f_{x_2x_2} & f_{x_2x_3} \\ f_{x_3x_1} & f_{x_3x_2} & f_{x_3x_3} \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 4 & 0 \\ 2 & 0 & 6 \end{bmatrix}$$

Definition: Third Leading Principal Minor

Just the determinant of $D^2 f$

$$D_3 = \begin{vmatrix} 2 & 2 & 2 \\ 2 & 4 & 0 \\ 2 & 0 & 6 \end{vmatrix} = 8$$

Definition: Second Leading Principal Minor

D_3 but you delete the last row/column

$$D_2 = \begin{vmatrix} 2 & 2 \\ 2 & 4 \end{vmatrix} = 4$$

Definition: First Leading Principal Minor

D_2 but you delete the last row/column

$$D_1 = \det([2]) = 2$$

Convexity test:

f is convex if and only if for all (x_1, x_2, x_3), we have $D_1, D_2, D_3 > 0$

Since $D_1 = 2, D_2 = 4, D_3 = 8 > 0$, we have that f is convex.

Note: In 2 dimensions, we only have D_2 (determinant) and D_1 (1×1 sub-determinant). In the previous example, you can check that $D_2 = 144 (x_1)^2 (x_2)^2 - 16$ is not always > 0, so f would not be convex.

Definition:

f is **concave** if $-f$ is convex

Warning: This is **NOT** the same thing as checking $D_1, D_2, D_3 < 0$ because $|-A| \neq |A|$

In fact, because $\left|-A\right| = (-1)^n \left|A\right|$ the correct thing would be to check

$$\begin{cases} (-1)^3 D_3 > 0 \\ (-1)^2 D_2 > 0 \\ (-1)^1 D_1 > 0 \end{cases} \Rightarrow \begin{cases} D_3 < 0 \\ D_2 > 0 \\ D_1 < 0 \end{cases}$$

In my opinion it's easier to just calculate $-f$ and check if that is convex.

Global Max/Min:

(1) If f is convex and (a, b) is a CP of f, then f has a global min at (a, b)

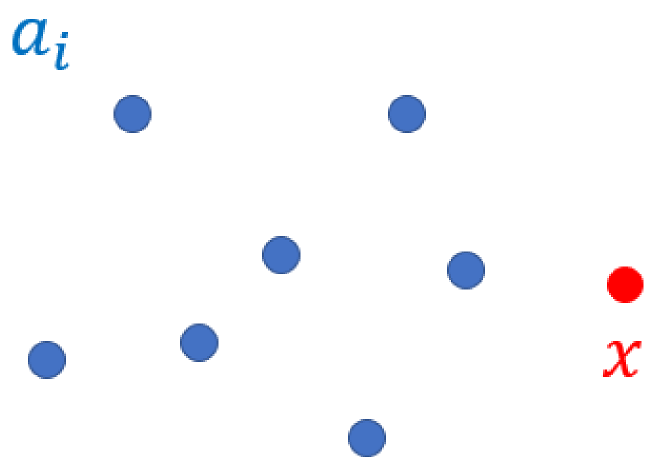
(2) If f is concave and (a, b) is a CP of f, then f has a global max at (a, b)

In our $f(x_1, x_2, x_3)$ example, can check that $(0, 0, 0)$ is a critical point of f. Since f is convex, it follows that f has a global min at $(0, 0, 0)$

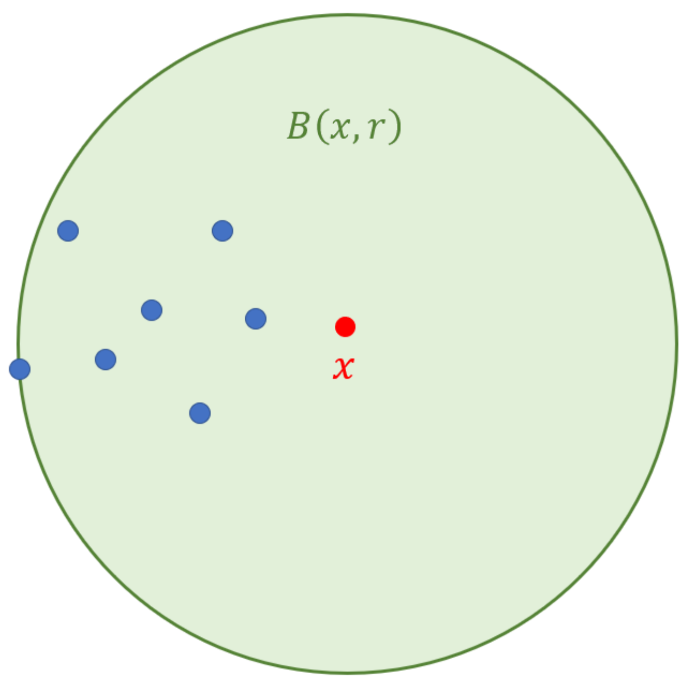
Here are three more applications of NLP. Mathematically, they are similar, but in practice they solve completely different problems.

3. Application 1: Clustering

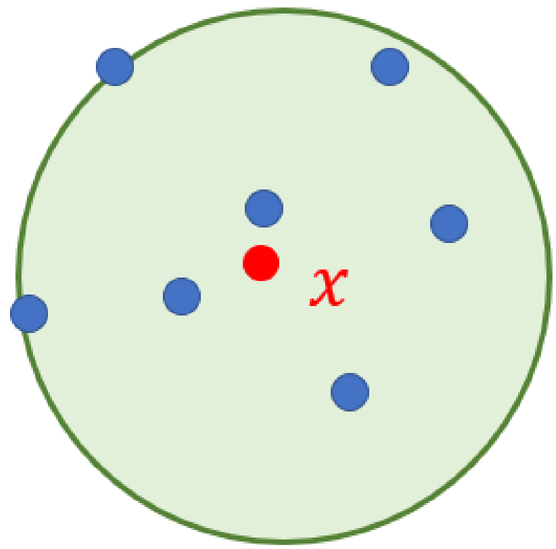
Suppose you have a collection $a_1, a_2, \dots, a_{100}$ of data points and a point x, think a pinned location.



Given x, consider the smallest (closed) ball $B(x,r)$ centered at x that goes through all the data points:



Notice r depends on x. In fact, the closer to are to the points, the smaller the radius:



Question: Where should you place x to get the smallest radius r?

We can express this as a NLP problem:

Notice $y \in B(x, r)$ if and only if $|y - x| \leq r$

So, in order for the data points a_i to be in $B(x, r)$ we need $|a_i - x| \leq r$

Therefore our minimization problem just becomes:

NLP Problem:

$$\begin{aligned} &\min r \\ \text{subject to } &|a_i - x| \leq r \text{ for all } i = 1, 2, \cdots, 100 \\ &x \in \mathbb{R}^n \end{aligned}$$

Explicit formula for r:

If you think about it, then r is just the biggest distance between x and all the data points, that is

$$r = \max\left\{|a_1 - x|, \cdots, |a_{100} - x|\right\}$$

By definition of max, we automatically have $|a_i - x| \leq r$ for all i and so our problem becomes:

Unconstrained NLP Problem:

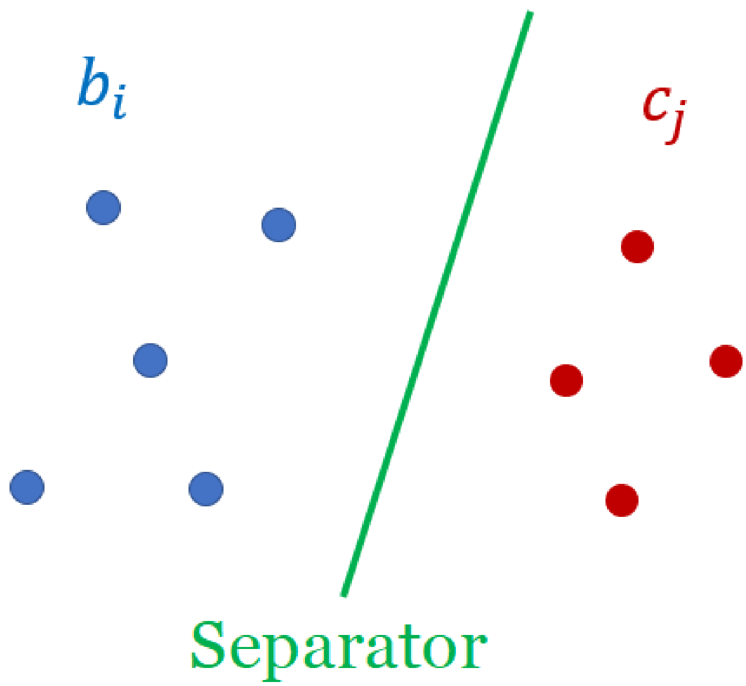
$$\begin{aligned} &\min_x \max_i \left\{|a_1 - x|, \ldots, |a_{100} - x|\right\} \\ &\text{subject to } x \in \mathbb{R}^n \end{aligned}$$

This is useful in applications to see if there are clusters of points, when r is small, or if the points spread out, when r is large

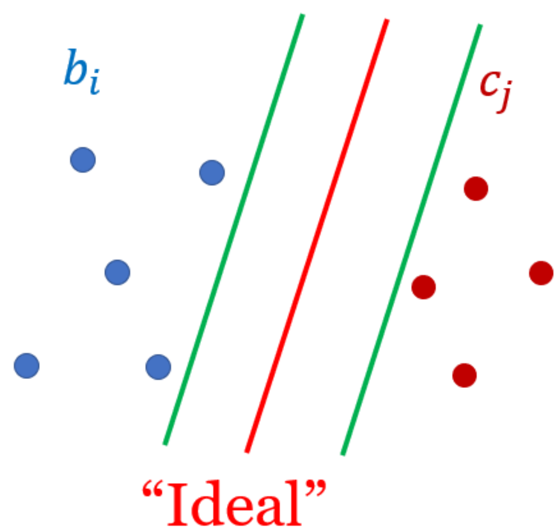
4. Application 2: Linear Separator

Suppose you have a collection $b_1, \ldots, b_{50}$ of blue points and $c_1, \ldots, c_{30}$ of red (crimson) points, say in 2 dimensions

Goal: Find a line that has all the blue points on one side, and all the red points on the other side



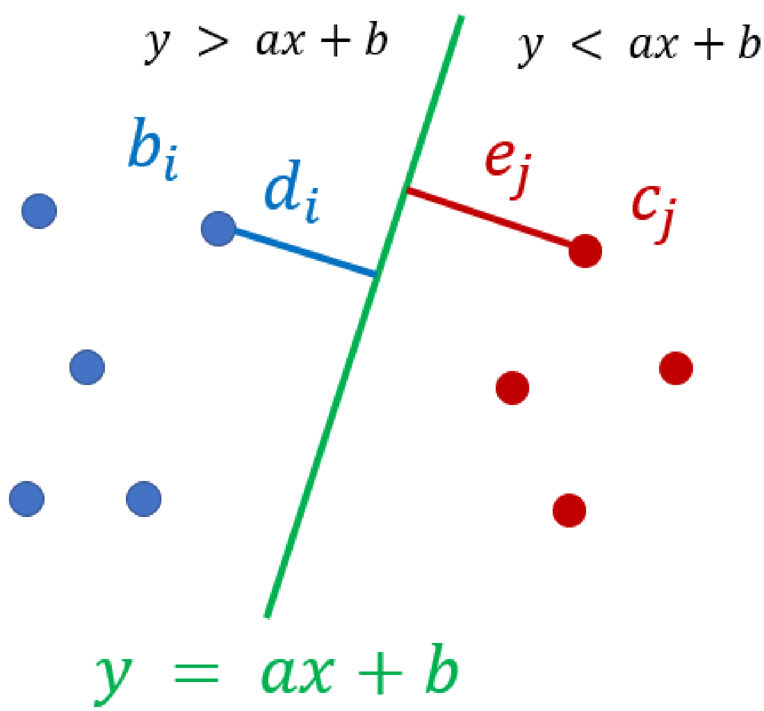
This line could not exist (if the points are not on one side), or there could be many such lines, such as follows:



We want to find a line that clearly separates the two. We can do so with distances:

$$d_i = \text{smallest distance between } b_i \text{ and line}$$
$$e_j = \text{smallest distance between } c_j \text{ and line}$$

Suppose our line as equation $y = ax + b$



We want the points to be as far away from the line as possible, as long as all the blue points are on one side of the line and the red ones are on the other side, so our problem becomes:

NLP Problem:

$$\max \sum_{i=1}^{50} d_i + \sum_{j=1}^{30} e_j$$
$$y > ax + b \text{ for all blue points}$$
$$y < ax + b \text{ for all red points}$$
$$a, b \text{ real}$$

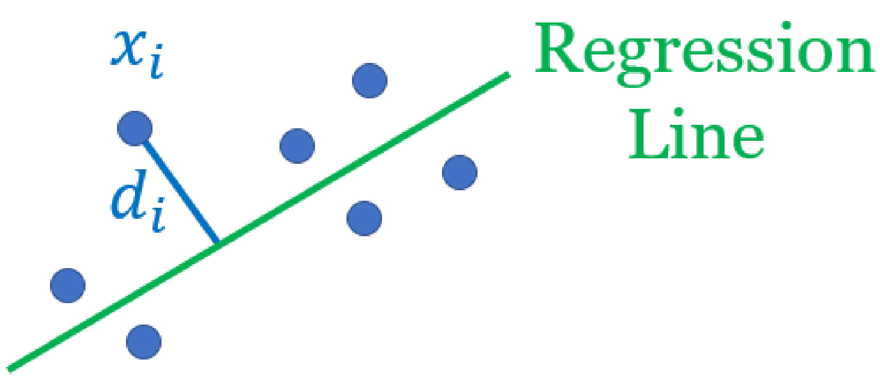
Can replace the sum by $\min \{d_i, e_j\}$ so it becomes a max-min problem

Note: Here a and b appear in the formulas for d_i and e_j, so it is a 2D max problem.

This is useful because once you have the line, and get a new data point, you can label that point as blue or red depending on which side of the line it is on.

5. Application 3: Regression

Given a collection of data points $x_1, \dots, x_{100}$, find the line that best fits the points



It's surprisingly the same problem, except instead of maximizing the distance, requiring the points to be far from the line, we are minimizing the distance, requiring the points to be close to the line.

$$d_i = \text{ smallest distance between } x_i \text{ and line}$$

Suppose our line is $y = ax + b$

Unconstrained NLP Problem:

$$\begin{aligned} \min d_1 + \cdots + d_{100} \\ \text{subject to } \ a, b \text{ real} \end{aligned}$$

Note: Can replace the sum with $\max d_i$ think minimizing the worst possible error.

Here a and b appear in the formulas for d_i so it is in fact a 2D min problem.