Intelligent Systems: Mathematics for AI
Extrema, Interpolation and Roots - Part I

Danesh Tarapore

Outline

- Finding extrema
- Interpolation, Taylor series
- Newton's method

Extrema

- In AI we often look for “best” solutions that maximise a goal function.

Extrema

- In AI we often look for "best" solutions that maximise a goal function.
- Many machine learning techniques are based on optimisation.
  - Neural networks: Finding optimal set of weights
  - Linear/Logistic regression: Finding optimal coefficients
  - SVM: Find optimal boundaries

Extrema

- In AI we often look for "best" solutions that maximise a goal function.
- Many machine learning techniques are based on optimisation.
  - Neural networks: Finding optimal set of weights
  - Linear/Logistic regression: Finding optimal coefficients
  - SVM: Find optimal boundaries
- Often this is done via numerical approximation techniques – mostly iteration schemes using Newton's method

Definitions

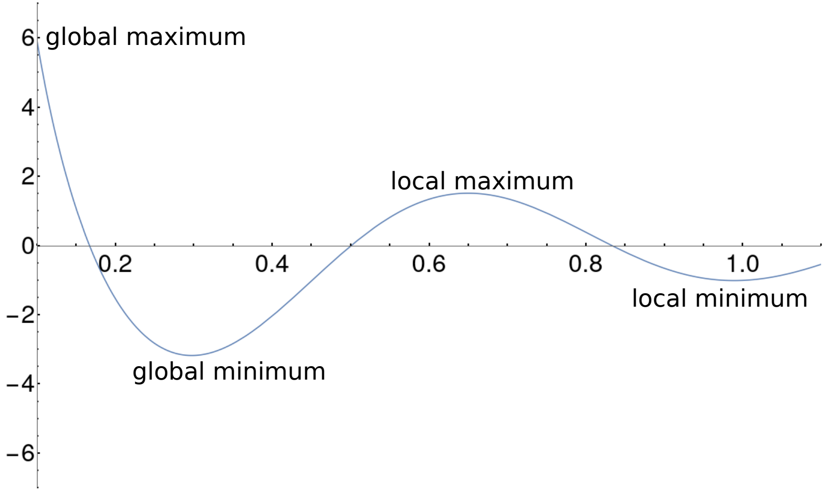
- Two types of extrema are of interest to us:
- Definition of global maximum: *A function f has a global maximum (minimum) at x^* if* $\forall x \neq x^* : f(x) \leq f(x^*)$ $(f(x) \geq f(x^*))$

Definitions

- Two types of extrema are of interest to us:
- Definition of global maximum: *A function f has a global maximum (minimum) at x^* if* $\forall x \neq x^* : f(x) \leq f(x^*)$ $(f(x) \geq f(x^*))$
- Definition of local optimum: *A function f has a local maximum (minimum) at x^* if for some* $\epsilon > 0$, $\forall x \in (x^* - \epsilon, x^* + \epsilon) : f(x) \leq f(x^*)$ $(f(x) \geq f(x^*))$

Extrema - example

- Plot the graph of
  $f(x) = \cos(3\pi x)/x$ for $x \in [0.1, 1.1]$



Fermat's Theorem (1)

- Let $f(x) \to R$ be a function, with domain $x \in A$.
- Suppose x_0 is a local extremum of f. If f is differentiable at x_0 then $f'(x_0) = 0$.

- Corollary: Global extrema of a function f on domain A occurs only at boundaries of A, non-differentiable points, and stationary points.

Fermat's Theorem (2)

- Why is this so?

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

- Suppose that x is a local maximum. By definition $f(x+h) \leq f(x)$. So $f(x+h) - f(x) \leq 0$.

Fermat's Theorem (3)

- $f(x+h) - f(x) \leq 0$.
- Consider $h > 0$:

$$\frac{f(x+h) - f(x)}{h} \leq 0 \implies \lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h}$$

Fermat's Theorem (3)

- $f(x+h) - f(x) \leq 0$.
- Consider $h > 0$:

$$\frac{f(x+h)-f(x)}{h} \leq 0 \Longrightarrow \lim_{h \to 0^+} \frac{f(x+h)-f(x)}{h}$$

- $= f'(x) \leq 0 \qquad (1)$

Fermat's Theorem (3)

- $f(x+h) - f(x) \leq 0.$
- Consider $h > 0$:

$$\frac{f(x+h) - f(x)}{h} \leq 0 \Longrightarrow \lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h}$$

- $= f'(x) \leq 0 \qquad (1)$
- Consider $h < 0$:

$$\frac{f(x+h) - f(x)}{h} \geq 0 \Longrightarrow \lim_{h \to 0^-} \frac{f(x+h) - f(x)}{h}$$

Fermat's Theorem (3)

- $f(x+h) - f(x) \leq 0.$
- Consider $h > 0$:

$$\frac{f(x+h) - f(x)}{h} \leq 0 \Longrightarrow \lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h}$$

- $= f'(x) \leq 0 \qquad (1)$
- Consider $h < 0$:

$$\frac{f(x+h) - f(x)}{h} \geq 0 \Longrightarrow \lim_{h \to 0^-} \frac{f(x+h) - f(x)}{h}$$

- $= f'(x) \geq 0 \qquad (2)$

Fermat's Theorem (3)

- $f(x+h) - f(x) \leq 0.$
- Consider $h > 0$:

$$\frac{f(x+h) - f(x)}{h} \leq 0 \implies \lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h}$$

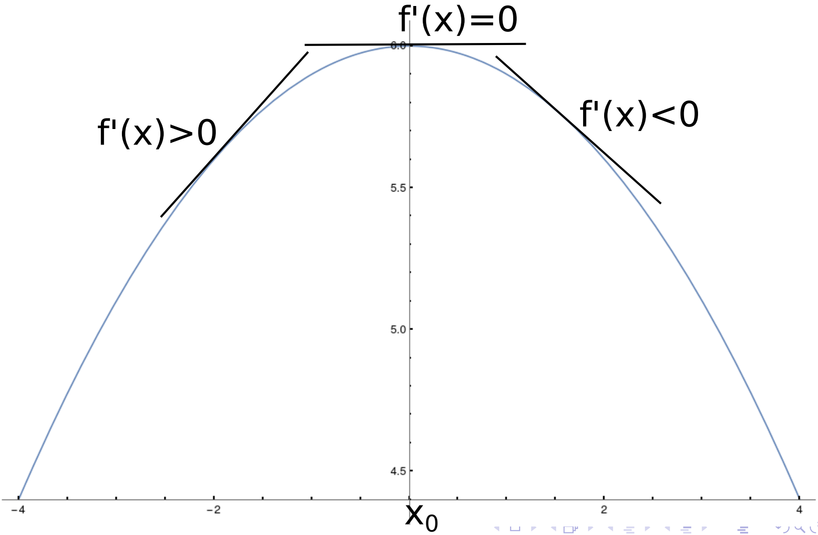
- $= f'(x) \leq 0 \qquad (1)$
- Consider $h < 0$:

$$\frac{f(x+h) - f(x)}{h} \geq 0 \implies \lim_{h \to 0^-} \frac{f(x+h) - f(x)}{h}$$

- $= f'(x) \geq 0 \qquad (2)$
- Combining Eqs. 1 and 2, $f'(x) = 0$.

Minimum or Maximum? (1)

- Suppose we found an extremum of f at x_0. Is it a maximum or a minimum?



Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$

Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$
  - From $+$ve to $-$ve: local maximum.

Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$
  - From $+$ve to $-$ve: local maximum.
  - From $-$ve to $+$ve: local minimum.

Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$
  - From $+$ve to $-$ve: local maximum.
  - From $-$ve to $+$ve: local minimum.
- Test this via higher order derivatives:

Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$
  - From $+$ve to $-$ve: local maximum.
  - From $-$ve to $+$ve: local minimum.
- Test this via higher order derivatives:
  - $f''(x) < 0$ – local maximum.

Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$
  - From $+$ve to $-$ve: local maximum.
  - From $-$ve to $+$ve: local minimum.
- Test this via higher order derivatives:
  - $f''(x) < 0$ – local maximum.
  - $f''(x) > 0$ – local minimum.

Minimum or Maximum? (2)

- So we need to look for a sign change in $f'(x)$
  - From $+$ve to $-$ve: local maximum.
  - From $-$ve to $+$ve: local minimum.
- Test this via higher order derivatives:
  - $f''(x) < 0$ – local maximum.
  - $f''(x) > 0$ – local minimum.
  - $f''(x) = 0$ – inconclusive, need to check higher order derivatives.

Why is this so? (1)

- Suppose x is a local extremum and we have $f''(x) < 0$.

Why is this so? (1)

- Suppose x is a local extremum and we have $f''(x) < 0$.
- Thus:

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} < 0$$

Why is this so? (1)

- Suppose x is a local extremum and we have $f''(x) < 0$.
- Thus:

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} < 0$$

- As $f'(x) = 0$ (local extremum), we get $\frac{f'(x+h)}{h} < 0$.

Why is this so? (2)

- There must be some environments where

$$\frac{f'(x+h)}{h} < 0$$

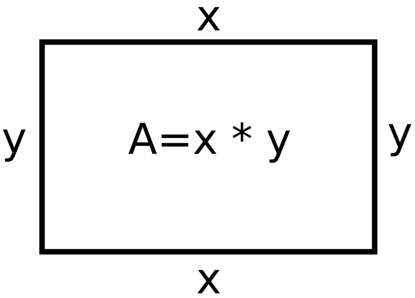
- What happens when h is $+$ve and when h is -ve?

$$f'(x) = \begin{cases} > 0 & \text{if } h < 0 \\ < 0 & \text{if } h > 0 \end{cases}$$

- We see that $f'(x)$ changes sign from "$+$" to "-", i.e., it is a local maxima.

Example (1)

- A farmer has a fence of length L at his disposal and wants to enclose the largest possible rectangular plot.
- What is the largest possible rectangular area he can enclose?



Constraint: Length of fence $L = 2x + 2y$

Example (1) - solution

- We are looking for max $A(x, y)$ subject to the constraint $L(x, y) = constant$.
- This reduces the problem to a uni-variate problem via,

$$L = 2x + 2y \longrightarrow y = L/2 - x$$
$$A = xy \longrightarrow A = x(L/2 - x) = L/(2x) - x^2$$

- Determining the local extrema:

$$\frac{dA}{dx} = 0 \longrightarrow L/2 - 2x = 0 \quad \boxed{x = L/4, A = L^2/16}$$

Example (1) - solution

- Local maxima or minima? Check 2^{nd} derivative.

$$\frac{d^2A}{dA^2} = \frac{d}{dx}\frac{dA}{dx} = \frac{d}{dx}(L/2 - 2x) = -2 < 0$$

- So it is a maxima!
- Global or local?

Example (1) - solution

- Local maxima or minima? Check 2^{nd} derivative.

$$\frac{d^2A}{dA^2} = \frac{d}{dx}\frac{dA}{dx} = \frac{d}{dx}(L/2 - 2x) = -2 < 0$$

- So it is a maxima!
- Global or local?
  - Check boundaries of domain $x \in [0, L/2]$ (boundaries from constrain $L(x, y)$)

Example (1) - solution

- Local maxima or minima? Check 2^{nd} derivative.

$$\frac{d^2 A}{dA^2} = \frac{d}{dx}\frac{dA}{dx} = \frac{d}{dx}(L/2 - 2x) = -2 < 0$$

- So it is a maxima!
- Global or local?
  - Check boundaries of domain $x \in [0, L/2]$ (boundaries from constrain $L(x, y)$)
  - $A(0) = 0$ and $A(L/2) = 0$, both $< L^2/16$

Example (1) - solution

- Local maxima or minima? Check 2^{nd} derivative.

$$\frac{d^2A}{dA^2} = \frac{d}{dx}\frac{dA}{dx} = \frac{d}{dx}(L/2 - 2x) = -2 < 0$$

- So it is a maxima!
- Global or local?
  - Check boundaries of domain $x \in [0, L/2]$ (boundaries from constrain $L(x, y)$)
  - $A(0) = 0$ and $A(L/2) = 0$, both $< L^2/16$
  - So a global maximum!

Example (2)

- Consider the function $f(x) = xe^{-x}$ over the interval $[0, \infty)$ and find all local/global minima/maxima
- The above function may be seen as an abstract model of payoff received when exploiting a resource which is subject to degradation
  - How much (x) would you harvest and how much do you "earn" from this?

Example (3)

- Consider the function $f(x) = x^2 \ln x$ over the interval $[1, 10]$ and find the extrema points and determine whether they are local/global maxima/minima.

Example (4)

- Find the local and global extrema of the function $f(x) = x^3 + 4x^2 - 18x$ over the interval $[-6, 6]$ and find the extrema points and determine whether they are local/global maxima/minima.

Summary

What is essential to remember:

- Optimisation idea.
- Fermat's theorem, criteria to distinguish global/local minima/maxima in a given function.

Next session

- Next math session we take a look at interpolation, Taylor series and the Newton's method to estimate the roots of an equation.