Modern Optimization Techniques

1. Theory

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute of Computer Science
University of Hildesheim, Germany

original slides by Lucas Rego Drumond (ISMLL)

Syllabus

Tue. 18.10.	(0)	0. Overview
		1. Theory
Tue. 25.10.	(1)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Tue. 1.11	(2)	2.1 Gradient Descent
Tue. 8.11.	(3)	2.2 Stochastic Gradient Descent
Tue. 15.11.	(4)	2.3 Newton's Method
Tue. 22.11.	(5)	2.3b Newton's Method (Part 2)
Tue. 29.11.	(6)	2.4 Subgradient Methods
Tue. 6.12.	(7)	2.5 Coordinate Descent
		3. Equality Constrained Optimization
Tue. 13.12.	(8)	3.1 Duality
Tue. 20.12.	(9)	3.2 Methods
	—	— Christmas Break —
		4. Inequality Constrained Optimization
Tue. 10.1.	(10)	4.1 Interior Point Methods
Tue. 17.1.	(11)	4.2 Cutting Plane Method
		5. Distributed Optimization
Tue. 24.1.	(11)	5.1 Alternating Direction Method of Multipliers
Tue. 31.1.	(12)	— Questions and Answers —

Outline

1. Introduction

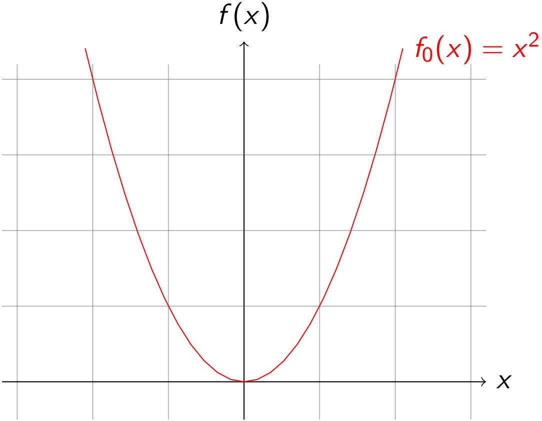
2. Convex Sets

3. Convex Functions

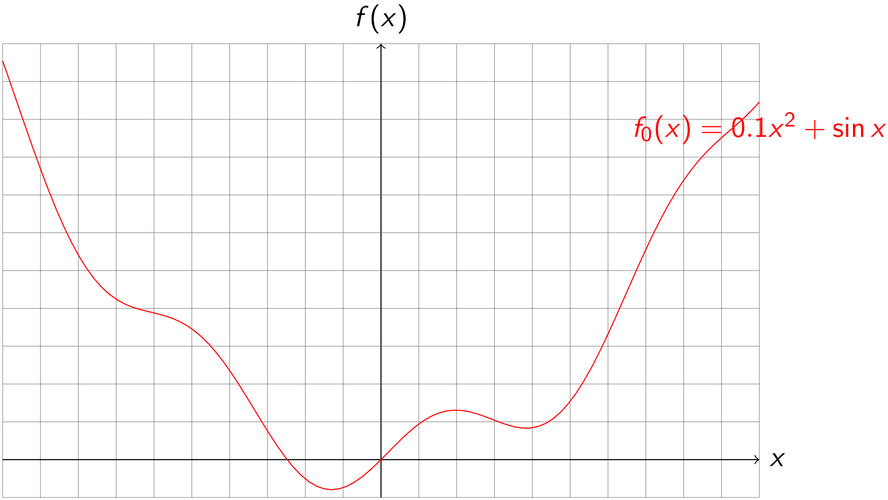
4. Optimization Problems

Outline

A convex function



A non-convex function



Convex Optimization Problem

An **optimization problem**

$$\begin{array}{rll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & Ax = b & \end{array}$$

is said to be convex if $f_0, \ldots f_m$ are convex

Convex Optimization Problem

An **optimization problem**

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

is said to be convex if $f_0, \ldots f_m$ are convex

How do we know if a function is convex or not?

Outline

Affine Sets

For any two points x_1, x_2 we can define the line through them as:

Affine Sets

For any two points x_1, x_2 we can define the line through them as:

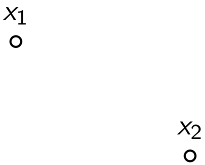
$$x = \theta x_1 + (1 - \theta) x_2 \qquad \theta \in \mathbb{R}$$

Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2 \qquad \theta \in \mathbb{R}$$

Example:

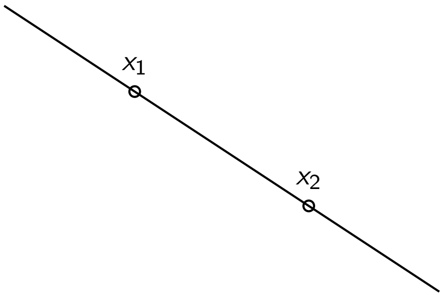


Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2 \qquad \theta \in \mathbb{R}$$

Example:

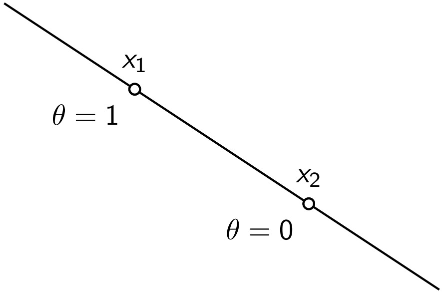


Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2 \qquad \theta \in \mathbb{R}$$

Example:

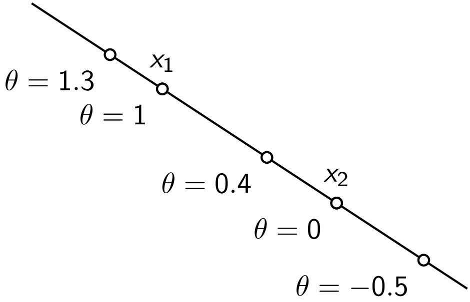


Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2 \qquad \theta \in \mathbb{R}$$

Example:



Affine Sets - Definition

An **affine set** is a set containing the line through any two distinct points in it

Affine Sets - Definition

An **affine set** is a set containing the line through any two distinct points in it

Examples:

- $\mathbb{R}^n$ for $n \in \mathbb{N}^+$
- Solution set of linear equations $\{x|Ax = b\}$

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2 \quad 0 \leq \theta \leq 1$$

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2 \quad 0 \leq \theta \leq 1$$

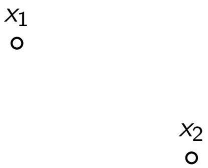
Example:

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2 \quad 0 \leq \theta \leq 1$$

Example:

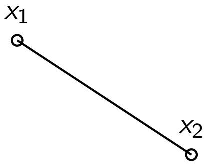


Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta)x_2 \quad 0 \leq \theta \leq 1$$

Example:

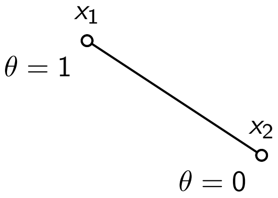


Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2 \quad 0 \leq \theta \leq 1$$

Example:

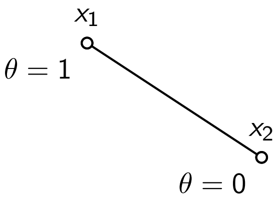


Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2 \quad 0 \leq \theta \leq 1$$

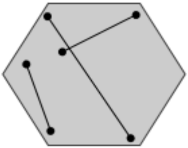
Example:

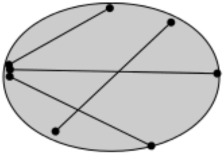


A **convex set** contains the line segment between any two points in the set

Convex Sets - Examples

Convex Sets:

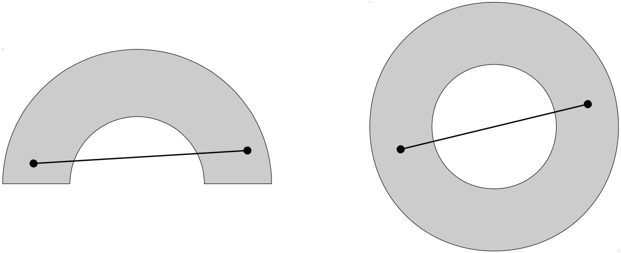




Convex Sets - Examples

Convex Sets:

Non-convex Sets:



Convex Combination and Convex Hull

(standard) simplex:

$$\Delta^N := \{\theta \in \mathbb{R}^N \mid \theta_n \geq 0, n = 1, \ldots, N; \sum_{n=1}^{N} \theta_n = 1\}$$

convex combination of some points $x_1, \ldots x_N \in \mathbb{R}^M$: any point x with

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_N x_N, \quad \theta \in \Delta^N$$

convex hull of a set $X \subseteq \mathbb{R}^M$ of points:

$$\text{conv}(X) := \{\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_N x_N \mid N \in \mathbb{N}, x_1, \ldots, x_N \in X, \theta \in \Delta^N\}$$

i.e., the set of all convex combinations of points in X.

Outline

Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **convex** iff:

Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **convex** iff:

- $\operatorname{dom} f = X$ is a convex set

Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **convex** iff:

- $\operatorname{dom} f = X$ is a convex set
- for all $x_1, x_2 \in \operatorname{dom} f$ and $0 \leq \theta \leq 1$ it satistfies

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

(the function is below any of its chords/secant segments.)

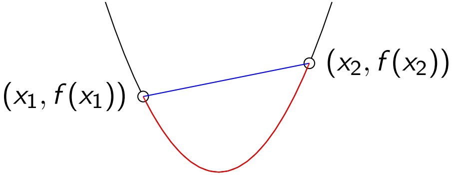
Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **convex** iff:

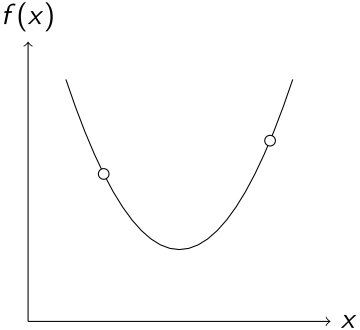
- $\operatorname{dom} f = X$ is a convex set
- for all $x_1, x_2 \in \operatorname{dom} f$ and $0 \leq \theta \leq 1$ it satistfies

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

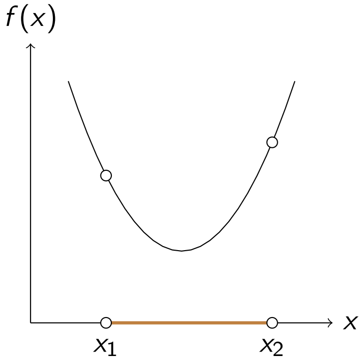
(the function is below any of its chords/secant segments.)



Convex functions

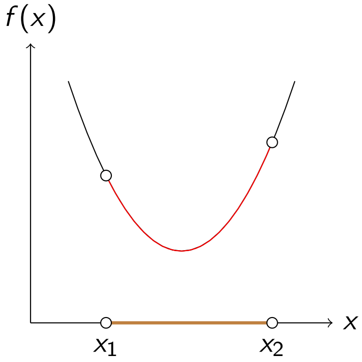


Convex functions



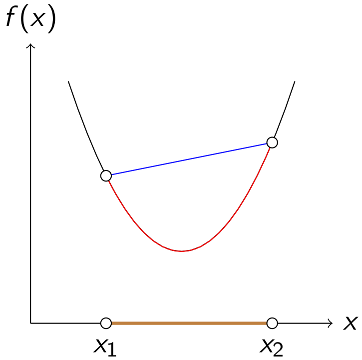
- $\theta x_1 + (1-\theta)x_2$

Convex functions



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

Convex functions



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

How are Convex Functions Related to Convex Sets?

epigraph of a function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$:

$$\text{epi}(f) := \{(x, y) \in X \times \mathbb{R} \mid y \geq f(x)\}$$

How are Convex Functions Related to Convex Sets?

epigraph of a function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$:

$$\text{epi}(f) := \{(x, y) \in X \times \mathbb{R} \mid y \geq f(x)\}$$

f is convex (as function) $\iff$ epi(f) is convex (as set).

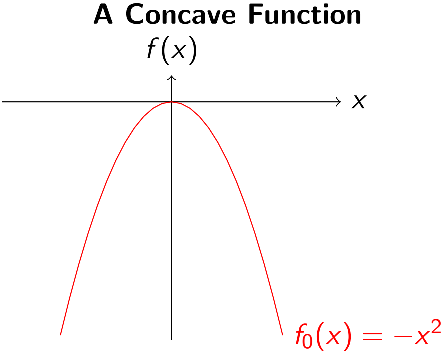
proof is straight-forward (try it!)

Concave Functions

A function f is called **concave** if $-f$ is convex

Concave Functions

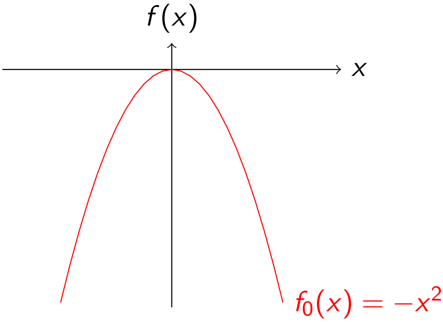
A function f is called **concave** if $-f$ is convex



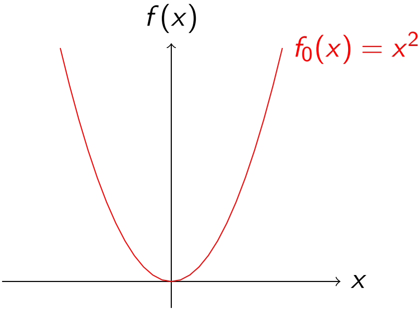
Concave Functions

A function f is called **concave** if $-f$ is convex

A Concave Function



A Convex Function



Strictly Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **strictly convex** if:

Strictly Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **strictly convex** if:

- dom f is a convex set

Strictly Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ is **strictly convex** if:

- dom f is a convex set
- for all $x_1, x_2 \in$ dom f, $x \neq y$ and $0 < \theta < 1$ it satistfies

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

Examples

Examples of Convex functions:

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\text{dom}\, f = \mathbb{R}$ and $a, b \in \mathbb{R}$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\operatorname{dom} f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\operatorname{dom} f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) = |x|^a$, with $\operatorname{dom} f = \mathbb{R}$ and $a \geq 1$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\text{dom}\, f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\text{dom}\, f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) = |x|^a$, with $\text{dom}\, f = \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) = x \log x$, with $\text{dom}\, f = \mathbb{R}^+$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\text{dom}\, f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\text{dom}\, f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) = |x|^a$, with $\text{dom}\, f = \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) = x \log x$, with $\text{dom}\, f = \mathbb{R}^+$

Examples of Concave Functions:

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\operatorname{dom} f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) = |x|^a$, with $\operatorname{dom} f = \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) = x \log x$, with $\operatorname{dom} f = \mathbb{R}^+$

Examples of Concave Functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\text{dom}\, f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\text{dom}\, f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) = |x|^a$, with $\text{dom}\, f = \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) = x \log x$, with $\text{dom}\, f = \mathbb{R}^+$

Examples of Concave Functions:

- affine: $f(x) = ax + b$, with $\text{dom}\, f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\text{dom}\, f = \mathbb{R}_0^+$ and $0 \leq a \leq 1$

Examples

Examples of Convex functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) = e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\operatorname{dom} f = \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) = |x|^a$, with $\operatorname{dom} f = \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) = x \log x$, with $\operatorname{dom} f = \mathbb{R}^+$

Examples of Concave Functions:

- affine: $f(x) = ax + b$, with $\operatorname{dom} f = \mathbb{R}$ and $a, b \in \mathbb{R}$
- powers: $f(x) = x^a$, with $\operatorname{dom} f = \mathbb{R}_0^+$ and $0 \leq a \leq 1$
- logarithm: $f(x) = \log x$, with $\operatorname{dom} f = \mathbb{R}^+$

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.
- For $\mathbf{x} \in \mathbb{R}^n$, $p \geq 1$:
  p-norms: $||\mathbf{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$,

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.
- For $\mathbf{x} \in \mathbb{R}^n$, $p \geq 1$:
  p-norms: $||\mathbf{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$,
- $||\mathbf{x}||_\infty = \max_k |x_k|$

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.
- For $\mathbf{x} \in \mathbb{R}^n$, $p \geq 1$:
  p-norms: $||\mathbf{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$,
- $||\mathbf{x}||_\infty = \max_k |x_k|$

Affine functions on vectors are also convex: $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$

1st-Order Condition

f is **differentiable** if $\operatorname{dom} f$ is open and the gradient

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)$$

exists everywhere.

1st-order condition: a differentiable function f is convex iff

1st-Order Condition

f is **differentiable** if $\operatorname{dom} f$ is open and the gradient

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)$$

exists everywhere.

1st-order condition: a differentiable function f is convex iff

- $\operatorname{dom} f$ is a convex set

1st-Order Condition

f is **differentiable** if dom f is open and the gradient

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)$$

exists everywhere.

1st-order condition: a differentiable function f is convex iff

- dom f is a convex set
- for all $\mathbf{x}, \mathbf{y} \in$ dom f

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

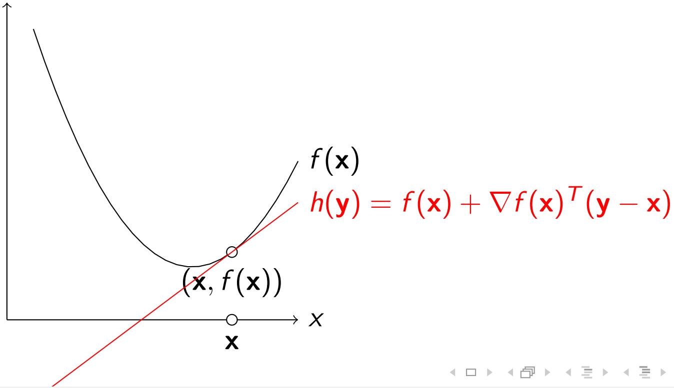
(the function is above any of its tangents.)

1st-Order Condition

1st-order condition: a differentiable function f is convex iff

- dom f is a convex set
- for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$

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



1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \rightarrow \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

“$\Rightarrow$” : $f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \rightarrow \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

$$\text{“} \Rightarrow \text{”} : f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$$

$$f(y) \geq \frac{f(x + t(y - x)) - f(x)}{t} + f(x) \overset{t \rightarrow 0^+}{\longrightarrow} \nabla f(x)^T(y - x) + f(x)$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

"$\Rightarrow$" : $f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$

$$f(y) \geq \frac{f(x + t(y - x)) - f(x)}{t} + f(x) \stackrel{t \to 0^+}{\longrightarrow} \nabla f(x)^T(y - x) + f(x)$$

"$\Leftarrow$" : Apply twice to $z := \theta x + (1 - \theta)y$

$$f(x) \geq f(z) + \nabla f(z)^T(x - z)$$

$$f(y) \geq f(z) + \nabla f(z)^T(y - z)$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

"$\Rightarrow$" : $f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$

$$f(y) \geq \frac{f(x + t(y-x)) - f(x)}{t} + f(x) \overset{t \to 0^+}{\longrightarrow} \nabla f(x)^T(y - x) + f(x)$$

"$\Leftarrow$" : Apply twice to $z := \theta x + (1 - \theta)y$

$$f(x) \geq f(z) + \nabla f(z)^T(x - z)$$

$$f(y) \geq f(z) + \nabla f(z)^T(y - z)$$

$$\rightsquigarrow \theta f(x) + (1 - \theta)f(y) \geq f(z) + \nabla f(z)^T(\theta x + (1 - \theta)y) - \nabla f(z)^T z$$

$$= f(z) + \nabla f(z)^T z - \nabla f(z)^T z = f(z) = f(\theta x + (1 - \theta)y)$$

1st-Order Condition / Strict Variant

strict 1st-order condition: a differentiable function f is strictly convex iff

- dom f is a convex set
- for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$

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

Global Minima

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

Consequence: Points x with $\nabla f(x) = 0$ are (equivalent) global minima.

- minima form a convex set
- if f is strictly convex: there is exactly one global minimum x^*.

2nd-Order Condition

f is **twice differentiable** if $\operatorname{dom} f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

exists everywhere.

2nd-order condition: a differentiable function f is convex iff

2nd-Order Condition

f is **twice differentiable** if $\operatorname{dom} f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

exists everywhere.

2nd-order condition: a differentiable function f is convex iff

- $\operatorname{dom} f$ is a convex set

2nd-Order Condition

f is **twice differentiable** if dom f is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

exists everywhere.

2nd-order condition: a differentiable function f is convex iff

- dom f is a convex set
- for all $\mathbf{x} \in \text{dom } f$

$$\nabla^2 f(\mathbf{x}) \succeq 0 \qquad \text{for all } \mathbf{x} \in \text{dom } f$$

2nd-Order Condition

f is **twice differentiable** if $\text{dom}\, f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

exists everywhere.

2nd-order condition: a differentiable function f is convex iff

- $\text{dom}\, f$ is a convex set
- for all $\mathbf{x} \in \text{dom}\, f$

$$\nabla^2 f(\mathbf{x}) \succeq 0 \qquad \text{for all } \mathbf{x} \in \text{dom}\, f$$

- if $\nabla^2 f(\mathbf{x}) \succ 0$ for all $\mathbf{x} \in \text{dom}\, f$, then f is strictly convex
  - the converse is not true,
    e.g., $f(x) = x^4$ is strictly convex, but has 0 derivative at 0.

Positive Semidefinite Matrices (A Reminder)

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semidefinite ($A \succeq 0$)**:

$$x^T A x \geq 0, \quad \forall x \in \mathbb{R}^n$$

Equivalent:

(i) all eigenvalues of A are ≥ 0.

(ii) $A = B^T B$ for some matrix B

Positive Semidefinite Matrices (A Reminder)

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is **positive semidefinite ($A \succeq 0$)**:

$$x^T A x \geq 0, \quad \forall x \in \mathbb{R}^n$$

Equivalent:

(i) all eigenvalues of A are ≥ 0.

(ii) $A = B^T B$ for some matrix B

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is **positive definite ($A \succ 0$)**:

$$x^T A x > 0, \quad \forall x \in \mathbb{R}^n \setminus \{0\}$$

Equivalent:

(i) all eigenvalues of A are > 0.

(ii) $A = B^T B$ for some nonsingular matrix B

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function
- If f can be obtained by applying those operations to a function, f is also convex

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function
- If f can be obtained by applying those operations to a function, f is also convex

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function
- If f can be obtained by applying those operations to a function, f is also convex

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex
- Example: $5x^2$ is convex since x^2 is convex

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function
- If f can be obtained by applying those operations to a function, f is also convex

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex
- Example: $5x^2$ is convex since x^2 is convex

Sum:

- if f_1 and f_2 are convex functions then $f_1 + f_2$ is convex

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function
- If f can be obtained by applying those operations to a function, f is also convex

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex
- Example: $5x^2$ is convex since x^2 is convex

Sum:

- if f_1 and f_2 are convex functions then $f_1 + f_2$ is convex
- Example: $f(x) = e^{3x} + x \log x$ with $\mathrm{dom}\, f = \mathbb{R}^+$ is convex since e^{3x} and $x \log x$ are convex

Recognizing Convex Functions

Composition with the affine function:

- if f is convex then $f(A\mathbf{x} + \mathbf{b})$ is convex

Recognizing Convex Functions

Composition with the affine function:

- if f is convex then $f(A\mathbf{x} + \mathbf{b})$ is convex
- Example: norm of an affine function $||A\mathbf{x} + \mathbf{b}||$

Recognizing Convex Functions

Composition with the affine function:

- if f is convex then $f(A\mathbf{x} + \mathbf{b})$ is convex
- Example: norm of an affine function $||A\mathbf{x} + \mathbf{b}||$

Pointwise Maximum:

- if $f_1, \ldots, f_m$ are convex functions then $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is convex

Recognizing Convex Functions

Composition with the affine function:

- if f is convex then $f(A\mathbf{x} + \mathbf{b})$ is convex
- Example: norm of an affine function $||A\mathbf{x} + \mathbf{b}||$

Pointwise Maximum:

- if $f_1, \ldots, f_m$ are convex functions then $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is convex
- Example: $f(\mathbf{x}) = \max_{i=1,\ldots,m}(a_i^T \mathbf{x} + b_i)$ is convex

Recognizing Convex Functions

Composition with scalar functions:

- if $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R} \rightarrow \mathbb{R}$ and

$$f(\mathbf{x}) = h(g(\mathbf{x}))$$

Recognizing Convex Functions

Composition with scalar functions:

- if $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R} \rightarrow \mathbb{R}$ and

$$f(\mathbf{x}) = h(g(\mathbf{x}))$$

- f is convex if:
  - g is convex, h is convex and h is nondecreasing *or*
  - g is concave, h is convex and h is nonincreasing

Recognizing Convex Functions

Composition with scalar functions:

- if $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R} \rightarrow \mathbb{R}$ and

$$f(\mathbf{x}) = h(g(\mathbf{x}))$$

- f is convex if:
  - g is convex, h is convex and h is nondecreasing *or*
  - g is concave, h is convex and h is nonincreasing
- Examples:
  - $e^{g(\mathbf{x})}$ is convex if g is convex
  - $\frac{1}{g(\mathbf{x})}$ is convex if g is concave and positive

Recognizing Convex Functions

There are many different ways to establish the convexity of a function:

- Apply the definition

Recognizing Convex Functions

There are many different ways to establish the convexity of a function:

- Apply the definition
- Show that $\nabla^2 f(\mathbf{x}) \succeq 0$ for twice differentiable functions

Recognizing Convex Functions

There are many different ways to establish the convexity of a function:

- Apply the definition
- Show that $\nabla^2 f(\mathbf{x}) \succeq 0$ for twice differentiable functions
- Show that f can be obtained from other convex functions by operations that preserve convexity

Outline

Optimization Problem

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, q \end{aligned}$$

- $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ is the **objective function**
- $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable
- $(f_i)_{i=1,\ldots,m} : \mathbb{R}^n \rightarrow \mathbb{R}$ are the inequality constraint functions
- $(h_i)_{i=1,\ldots,q} : \mathbb{R}^n \rightarrow \mathbb{R}$ are the equality constraint functions

Convex Optimization Problem

An **optimization problem**

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, q \end{aligned}$$

is said to be convex if $f_0, \ldots f_p$ are convex and $h_1, \ldots, h_q$ are **affine**:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p \\ & A\mathbf{x} = \mathbf{b} \end{aligned}$$

Practical Example: Household Spending

Suppose we have the following data about different households:

- Number of workers in the household (a_1)
- Household composition (a_2)
- Region (a_3)
- Gross normal weekly household income (a_4)
- **Weekly household spending** (y)

Practical Example: Household Spending

Suppose we have the following data about different households:

- Number of workers in the household (a_1)
- Household composition (a_2)
- Region (a_3)
- Gross normal weekly household income (a_4)
- **Weekly household spending** (y)

We want to create a model of the weekly household spending

Practical Example: Household Spending

If we have data about m households, we can represent it as:

$$A = \begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Practical Example: Household Spending

If we have data about m households, we can represent it as:

$$A = \begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

We can model the household consumption is a linear combination of the household features with parameters β:

$$\begin{aligned} \hat{y}_i =& \beta^T \mathbf{a}_i = \beta_0 1 + \beta_1 a_{i,1} + \beta_2 a_{i,2} + \beta_3 a_{i,3} + \beta_4 a_{i,4} \\ \mathbf{a}_i :=& A_{i,.} \end{aligned}$$

Practical Example: Household Spending

We have:

$$\begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \approx \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Practical Example: Household Spending

We have:

$$\begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \approx \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

We want to find parameters β such that the measured error of the predictions is minimal:

$$\sum_{i=1}^{m} (\beta^T \mathbf{a_i} - y_i)^2 = ||A\beta - \mathbf{y}||_2^2$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

$$\frac{d}{d\beta}(A\beta - \mathbf{y})^T(A\beta - \mathbf{y}) = 2A^T(A\beta - \mathbf{y})$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

$$\frac{d}{d\beta}(A\beta - \mathbf{y})^T(A\beta - \mathbf{y}) = 2A^T(A\beta - \mathbf{y})$$

$$2A^T(A\beta - \mathbf{y}) = 0$$

The Least Squares Problem

$$\text{minimize} \qquad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

$$\frac{d}{d\beta}(A\beta - \mathbf{y})^T(A\beta - \mathbf{y}) = 2A^T(A\beta - \mathbf{y})$$

$$2A^T(A\beta - \mathbf{y}) = 0$$
$$A^TA\beta - A^T\mathbf{y} = 0$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

$$\frac{d}{d\beta}(A\beta - \mathbf{y})^T(A\beta - \mathbf{y}) = 2A^T(A\beta - \mathbf{y})$$

$$\begin{aligned} 2A^T(A\beta - \mathbf{y}) &= 0 \\ A^T A\beta - A^T\mathbf{y} &= 0 \\ A^T A\beta &= A^T\mathbf{y} \end{aligned}$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

$$\frac{d}{d\beta}(A\beta - \mathbf{y})^T(A\beta - \mathbf{y}) = 2A^T(A\beta - \mathbf{y})$$

$$\begin{aligned} 2A^T(A\beta - \mathbf{y}) &= 0 \\ A^TA\beta - A^T\mathbf{y} &= 0 \\ A^TA\beta &= A^T\mathbf{y} \\ \beta &= (A^TA)^{-1}A^T\mathbf{y} \end{aligned}$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

- Convex Problem!
- Analytical solution: $\beta^* = (A^T A)^{-1} A^T \mathbf{y}$
- Often applied for data fitting
- $A\beta - \mathbf{y}$ is usually called the residual or error
- Extensions such as regularized least squares

Practical Example: Household Location

Suppose we have the following data about different households:

- Number of workers in the household (a_1)
- Household composition (a_2)
- Weekly household spending (a_3)
- Gross normal weekly household income (a_4)
- **Region** (y): north $y = 1$ or south $y = 0$

Practical Example: Household Location

Suppose we have the following data about different households:

- Number of workers in the household (a_1)
- Household composition (a_2)
- Weekly household spending (a_3)
- Gross normal weekly household income (a_4)
- **Region** (y): north $y = 1$ or south $y = 0$

We want to create a model of the location of the household

Practical Example: Household Location

If we have data about m households, we can represent it as:

$$A = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,n} \\ 1 & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,2} & \dots & a_{m,n} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Practical Example: Household Location

If we have data about m households, we can represent it as:

$$A = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,n} \\ 1 & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,2} & \dots & a_{m,n} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

We can model the probability of the household location to be north ($y = 1$) as a linear combination of the household features with parameters β:

$$\hat{y}_i = \sigma(\beta^T \mathbf{a_i}) = \sigma(\beta_0 1 + \beta_1 a_{i,1} + \beta_2 a_{i,2} + \beta_3 a_{i,3} + \beta_4 a_{i,4})$$

where: $\sigma(x) := \frac{1}{1+e^{-x}}$ (logistic function)

Logistic Regression

The logistic regression learning problem is

$$\text{maximize} \quad \sum_{i=1}^{m} y_i \log \sigma(\beta^T \mathbf{a_i}) + (1 - y_i) \log(1 - \sigma(\beta^T \mathbf{a_i}))$$

$$A = \begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Linear Programming

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{c}^T \mathbf{x} \\
\text{subject to} \quad & \mathbf{a}_i^T \mathbf{x} \leq b_i \quad i = 1, \ldots, m
\end{aligned}
$$

Linear Programming

$$\begin{array}{rll} \text{minimize} & \mathbf{c}^T\mathbf{x} & \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i & i = 1, \ldots, m \end{array}$$

- No simple analytical solution

Linear Programming

$$\begin{array}{rll} \text{minimize} & \mathbf{c}^T\mathbf{x} & \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i & i = 1, \ldots, m \end{array}$$

- No simple analytical solution
- There are reliable algorithms available:

Linear Programming

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{a}_i^T \mathbf{x} \leq b_i \quad i = 1, \ldots, m \end{aligned}$$

- No simple analytical solution
- There are reliable algorithms available:
  - Simplex

Linear Programming

$$
\begin{array}{ll}
\text{minimize} & \mathbf{c}^T\mathbf{x} \\
\text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i \quad i = 1, \ldots, m
\end{array}
$$

- No simple analytical solution
- There are reliable algorithms available:
  - Simplex
  - Interior Points Method

Summary (1/2)

- **Convex sets** are closed under line segments (convex combinations).
- **Convex functions** are defined on a convex domain and
  - are below any of their chords / secants (definition)
  - are globally above their tangents (1st-order condition)
  - have a positive definite Hessian (2nd-order condition)
- For convex functions, points with vanishing gradients are (equivalent) **global minima**.
- Operations that preserve convexity:
  - scaling with a nonnegative constant
  - sums
  - pointwise maximum
  - composition with an affine function
  - composition with a nondecreasing convex scalar function
  - composition of a nonincreasing convex scalar function with a concave function
    - esp. $-g$ for a concave g

Summary (2/2)

- General optimization problems consist of
  - an objective function,
  - inequality constraints and
  - equality constraints.

- **Convex optimization problems** have
  - a convex objective function,
  - convex inequality constraints and
  - affine equality constraints.

- Examples for convex optimization problems:
  - linear regression / least squares
  - linear classification / logistic regression
  - linear programming
  - quadratic programming

Further Readings

- Convex sets:
  - Boyd and Vandenberghe [2004], chapter 2, esp. 2.1
  - see also ch. 2.2 and 2.3

- Convex functions:
  - Boyd and Vandenberghe [2004], chapter 3, esp. 3.1.1–7, 3.2.1–5

- Convex optimization:
  - Boyd and Vandenberghe [2004], chapter 4, esp. 4.1–3
  - see also ch. 4.4

References I

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge Univ Press, 2004.