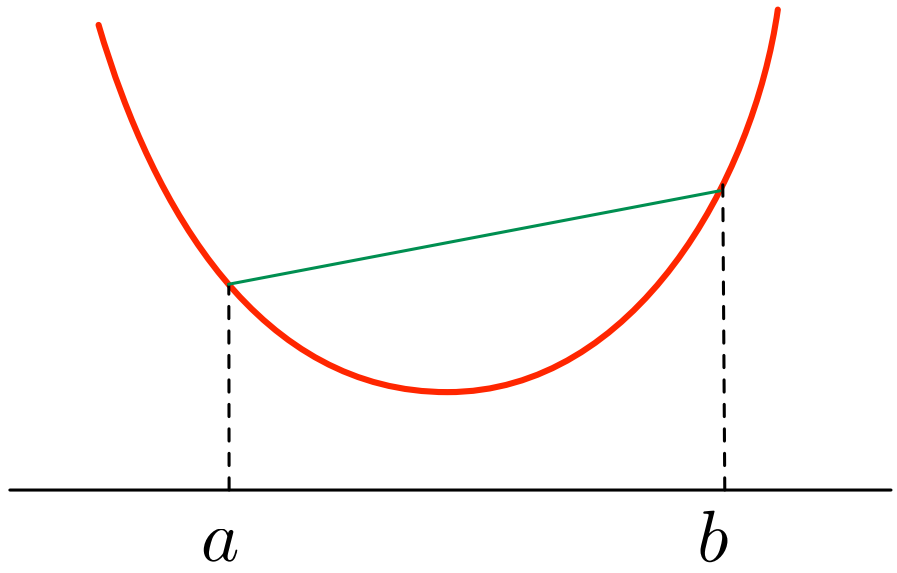
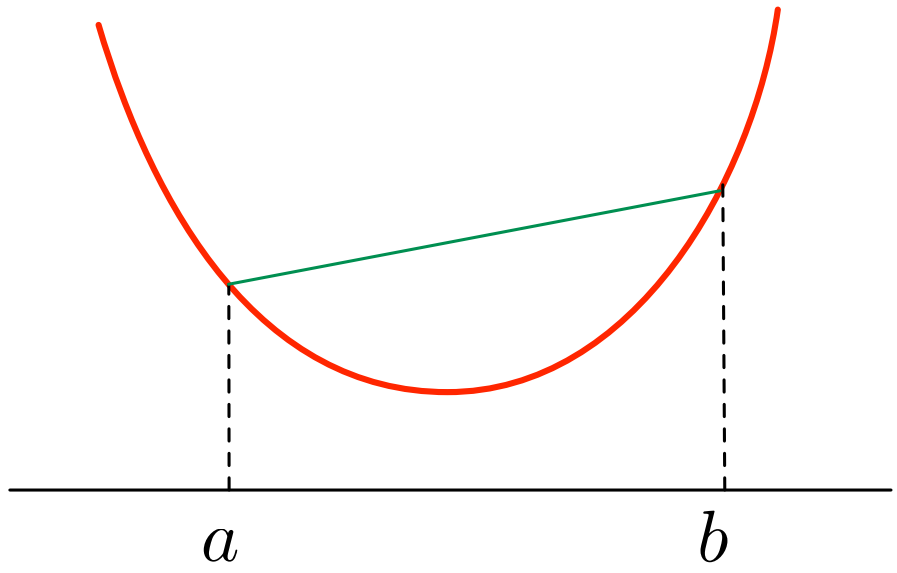
Convexity

Instructor: Taylor Berg-Kirkpatrick
Slides: Sanjoy Dasgupta
Course website:
http://cseweb.ucsd.edu/classes/wi19/cse151-b/

Convexity



Convexity

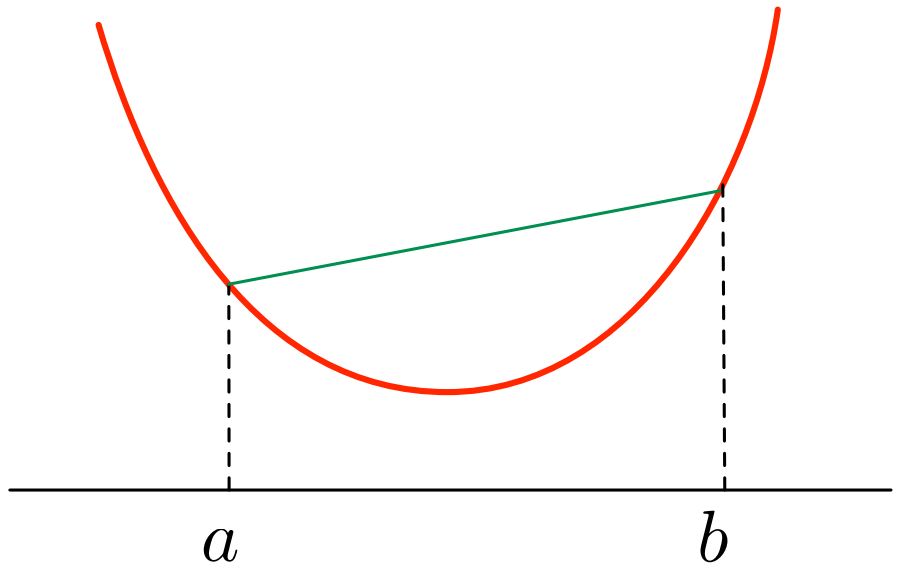


A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if for all $a, b \in \mathbb{R}^d$ and $0 < \theta < 1$,

$$f(\theta a + (1-\theta)b) \;\leq\; \theta f(a) + (1-\theta)f(b).$$

It is **strictly convex** if strict inequality holds for all $a \neq b$.

Convexity



A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if for all $a, b \in \mathbb{R}^d$ and $0 < \theta < 1$,

$$f(\theta a + (1-\theta)b) \;\leq\; \theta f(a) + (1-\theta)f(b).$$

It is **strictly convex** if strict inequality holds for all $a \neq b$.

f is **concave** $\Leftrightarrow$ $-f$ is convex

Checking convexity

A function on one variable, $f : \mathbb{R} \to \mathbb{R}$, is convex if its second derivative is ≥ 0 everywhere.

Checking convexity

A function on one variable, $f : \mathbb{R} \to \mathbb{R}$, is convex if its second derivative is ≥ 0 everywhere.

Example: $f(z) = z^2$

First and second derivatives of multivariate functions

For a function $f : \mathbb{R}^d \to \mathbb{R}$,

- the first derivative is a vector with d entries:

$$\nabla f(z) = \begin{pmatrix} \frac{\partial f}{\partial z_1} \\ \vdots \\ \frac{\partial f}{\partial z_d} \end{pmatrix}$$

- the second derivative is a $d \times d$ matrix, the **Hessian** $H(z)$:

$$H_{jk} = \frac{\partial^2 f}{\partial z_j \partial z_k}$$

Example

Find the second derivative matrix of $f(z) = \|z\|^2$.

Example

Find the second derivative matrix of $f(z) = \|z\|^2$.

$$f(z) = \sum_{i=1}^{d} z_i^2$$

Example

Find the second derivative matrix of $f(z) = \|z\|^2$.

$$f(z) = \sum_{i=1}^{d} z_i^2$$

$$\nabla f(z) = \begin{bmatrix} 2z_1 \\ \vdots \\ 2z_d \end{bmatrix}$$

Example

Find the second derivative matrix of $f(z) = \|z\|^2$.

$$f(z) = \sum_{i=1}^{d} z_i^2$$

$$\nabla f(z) = \begin{bmatrix} 2z_1 \\ \vdots \\ 2z_d \end{bmatrix}$$

$$\nabla^2 f(z) = 2I$$

Second-derivative test for convexity

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if its matrix of second derivatives is **positive semidefinite** everywhere.

Second-derivative test for convexity

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if its matrix of second derivatives is **positive semidefinite** everywhere.

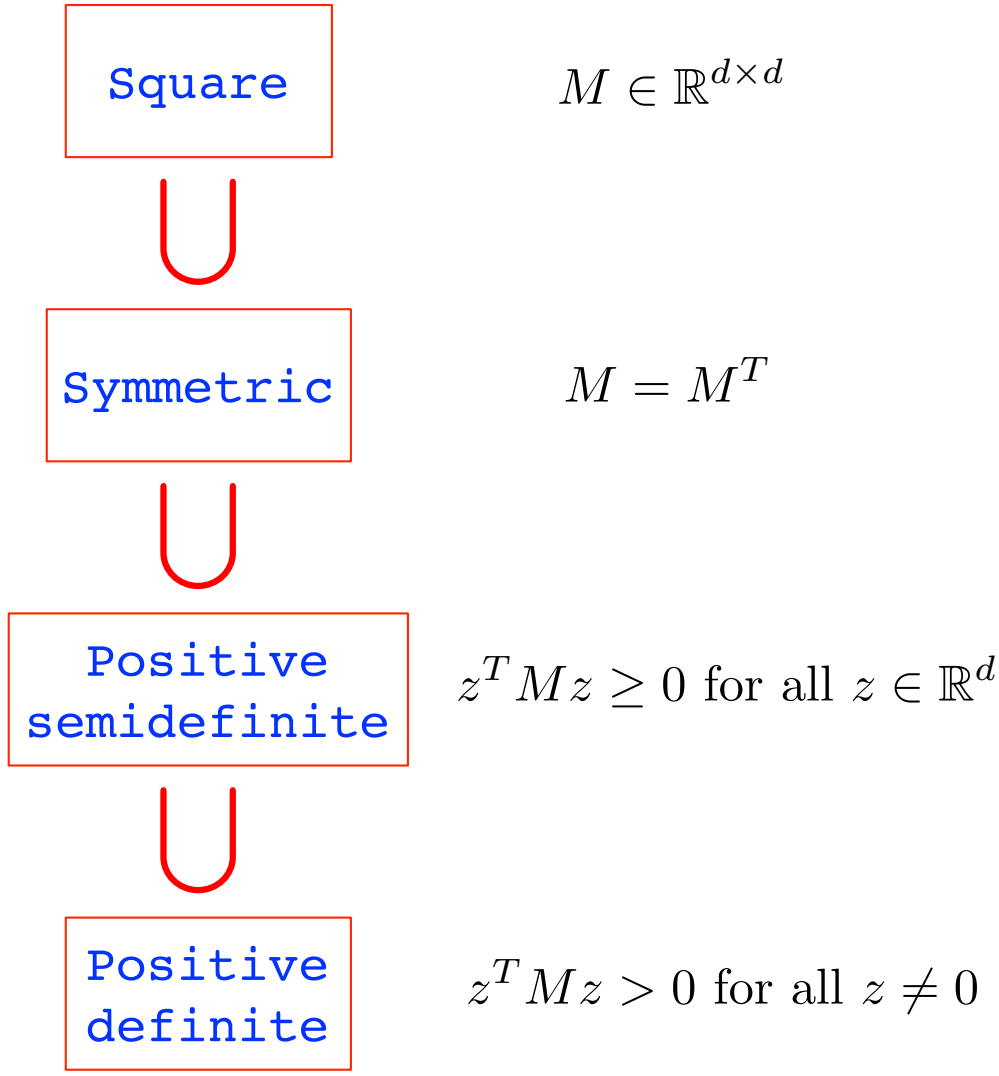
Recall: every **square** matrix M encodes a **quadratic function**:

$$x \mapsto x^T M x = \sum_{i,j=1}^{d} M_{ij} x_i x_j$$

(M is a $d \times d$ matrix and x is a vector in $\mathbb{R}^d$)

Sometimes $x^T M x$ is always ≥ 0, no matter what x you plug in.

A hierarchy of square matrices



A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

PSD or not?

- $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

PSD or not?

- $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

$$\begin{aligned} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} \\ &= x_1^2 + 2x_1x_2 + x_2^2 \\ &= (x_1 + x_2)^2 \end{aligned}$$

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

PSD or not?

- $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

$$\begin{aligned} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} \\ &= x_1^2 + 2x_1x_2 + x_2^2 \\ &= (x_1 + x_2)^2 \end{aligned}$$

- $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

PSD or not?

- $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

$$\begin{aligned} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} \\ &= x_1^2 + 2x_1x_2 + x_2^2 \\ &= (x_1 + x_2)^2 \end{aligned}$$

- $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

$$\begin{aligned} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= x_1^2 + 4x_1x_2 + x_2^2 \\ &= (x_1 + x_2)^2 + 2x_1x_2 \end{aligned}$$

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

When is a diagonal matrix PSD?

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

When is a diagonal matrix PSD?

$$A = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{bmatrix}$$

$$x^T A x = a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2$$

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

If M, N are of the same size and PSD, must $M + N$ be PSD?

A symmetric matrix M is **positive semidefinite (psd)** if:

$$z^T M z \geq 0 \quad \text{for all vectors } z$$

If M, N are of the same size and PSD, must $M + N$ be PSD?

$$x^T (M + N) x = x^T M x + x^T N x$$

Checking if a matrix is PSD

A matrix M is PSD if and only if it can be written as $M = UU^T$ for some matrix U.

Checking if a matrix is PSD

A matrix M is PSD if and only if it can be written as $M = UU^T$ for some matrix U.

Quick check: say $U \in \mathbb{R}^{r \times d}$ and $M = UU^T$.

1. M is square.
2. M is symmetric.
3. Pick any $z \in \mathbb{R}^r$. Then

$$\begin{aligned} z^T M z \;=\; z^T U U^T z \;&=\; (z^T U)(U^T z) \\ &=\; (U^T z)^T (U^T z) \\ &=\; \|U^T z\|^2 \;\geq\; 0. \end{aligned}$$

Checking if a matrix is PSD

A matrix M is PSD if and only if it can be written as $M = UU^T$ for some matrix U.

Quick check: say $U \in \mathbb{R}^{r \times d}$ and $M = UU^T$.

1. M is square.
2. M is symmetric.
3. Pick any $z \in \mathbb{R}^r$. Then

$$\begin{aligned} z^T M z \;=\; z^T U U^T z \;&=\; (z^T U)(U^T z) \\ &=\; (U^T z)^T (U^T z) \\ &=\; \|U^T z\|^2 \;\geq\; 0. \end{aligned}$$

Another useful fact: any covariance matrix is PSD.

Another useful fact: any covariance matrix is PSD.

$$\begin{aligned}\Sigma &= \frac{1}{N}(X - M)^T(X - M) \\ &= \left(\frac{1}{\sqrt{N}}(X - M)\right)^T \left(\frac{1}{\sqrt{N}}(X - M)\right)\end{aligned}$$

where $M \in \mathbb{R}_{N \times d}$ and

$$M = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}$$

Second-derivative test for convexity

A function (of several variables) is convex if its second-derivative matrix is **positive semidefinite** everywhere.

More formally:
Suppose that for $f : \mathbb{R}^d \to \mathbb{R}$, the second partial derivatives exist everywhere and are continuous functions of z. Then:

1. $H(z)$ is a symmetric matrix
2. f is convex $\Leftrightarrow$ $H(z)$ is positive semidefinite for all $z \in \mathbb{R}^d$

Example

Is $f(x) = \|x\|^2$ convex?

Example

Is $f(x) = \|x\|^2$ convex?

- Recall:

$$\nabla^2 f(x) = 2I$$

Example

Is $f(x) = \|x\|^2$ convex?

- Recall:

$$\nabla^2 f(x) = 2I$$

- This is a diagonal matrix with all positive entries along the diagonal.

Fix any vector $u \in \mathbb{R}^d$. Is this function $f : \mathbb{R}^d \to \mathbb{R}$ convex?

$$f(z) = (u \cdot z)^2$$

Fix any vector $u \in \mathbb{R}^d$. Is this function $f : \mathbb{R}^d \to \mathbb{R}$ convex?

$$f(z) = (u \cdot z)^2$$

$$\nabla f(z) = 2(u \cdot z)u$$

$$\nabla^2 f(z) = 2 \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_d \\ u_1 u_2 & u_2^2 & \cdots & u_2 u_d \\ \vdots & \vdots & \vdots & \vdots \\ u_1 u_d & u_2 u_d & \cdots & u_d^2 \end{bmatrix}$$

$$= 2uu^T$$

Least-squares regression

Recall loss function: for data points $(x^{(i)}, y^{(i)}) \in \mathbb{R}^d \times \mathbb{R}$,

$$L(w) = \sum_{i=1}^{n} (y^{(i)} - (w \cdot x^{(i)}))^2$$

Logistic regression

Recall loss function: for data $(x^{(i)}, y^{(i)}) \in \mathbb{R}^d \times \{-1, +1\}$,

$$L(w) = \sum_{i=1}^{n} \ln(1 + e^{-y^{(i)}(w \cdot x^{(i)})})$$

Logistic regression

Recall loss function: for data $(x^{(i)}, y^{(i)}) \in \mathbb{R}^d \times \{-1, +1\}$,

$$L(w) = \sum_{i=1}^{n} \ln(1 + e^{-y^{(i)}(w \cdot x^{(i)})})$$

We earlier found the first derivative:

$$\frac{\partial L}{\partial w_j} = -\sum_{i=1}^{n} y^{(i)} x_j^{(i)} \frac{1}{1 + e^{y^{(i)}(w \cdot x^{(i)})}}.$$

Logistic regression, cont'd

Second derivative: the (j, k) entry of the Hessian $H(w)$ is

$$\frac{\partial L}{\partial w_k \partial w_j} = \sum_{i=1}^{n} \left(x_j^{(i)} x_k^{(i)} \frac{1}{1 + e^{w \cdot x^{(i)}}} \frac{1}{1 + e^{-w \cdot x^{(i)}}} \right)$$

Logistic regression, cont'd

Second derivative: the (j, k) entry of the Hessian $H(w)$ is

$$\frac{\partial L}{\partial w_k \partial w_j} = \sum_{i=1}^{n} \left(x_j^{(i)} x_k^{(i)} \frac{1}{1 + e^{w \cdot x^{(i)}}} \frac{1}{1 + e^{-w \cdot x^{(i)}}} \right)$$

This is $u_j \cdot u_k$, where vectors $u_1, \ldots, u_d \in \mathbb{R}^n$ are defined as follows:

u_j has ith coordinate $x_j^{(i)} \sqrt{\frac{1}{(1 + e^{w \cdot x^{(i)}})(1 + e^{-w \cdot x^{(i)}})}}$

Logistic regression, cont'd

Second derivative: the (j, k) entry of the Hessian $H(w)$ is

$$\frac{\partial L}{\partial w_k \partial w_j} = \sum_{i=1}^{n} \left(x_j^{(i)} x_k^{(i)} \frac{1}{1 + e^{w \cdot x^{(i)}}} \frac{1}{1 + e^{-w \cdot x^{(i)}}} \right)$$

This is $u_j \cdot u_k$, where vectors $u_1, \ldots, u_d \in \mathbb{R}^n$ are defined as follows:

u_j has ith coordinate $x_j^{(i)} \sqrt{\frac{1}{(1 + e^{w \cdot x^{(i)}})(1 + e^{-w \cdot x^{(i)}})}}$

Therefore $H(w) = UU^T$, where U is the matrix with rows u_j.
Convex!