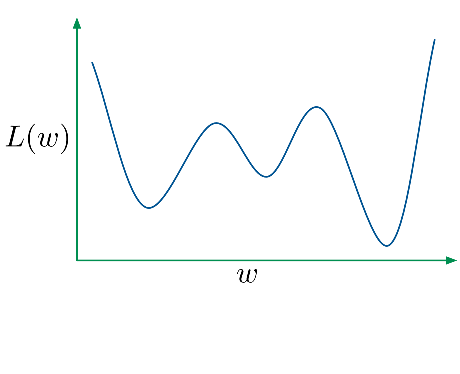
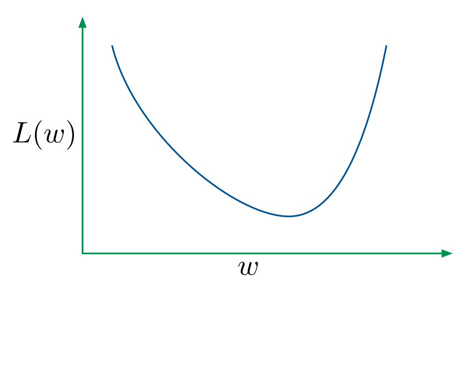
Convexity

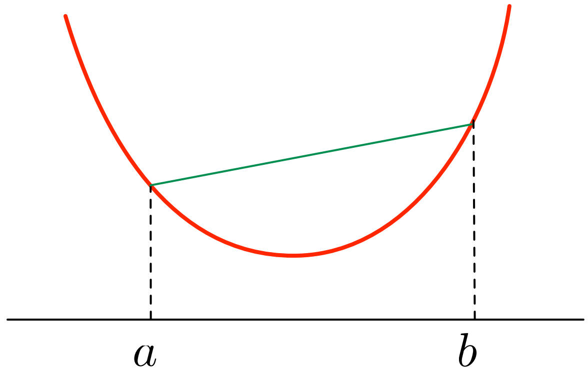
Is our loss function convex?



versus



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 for functions of one variable

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

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

Checking convexity

Function of one variable

$$F : \mathbb{R} \to \mathbb{R}$$

- Value: number
- Derivative: number
- Second derivative: number

Convex if second derivative is always ≥ 0

Function of d variables

$$F : \mathbb{R}^d \to \mathbb{R}$$

- Value: number
- Derivative: d-dimensional vector
- Second derivative: $d \times d$ matrix

Convex if second derivative matrix is always positive semidefinite

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 $f(z) = \|z\|^2$.

When is a square matrix "positive"?

- A superficial notion: when all its entries are positive
- A deeper notion: **when the quadratic function defined by it is always positive**

Example: $M = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

Positive semidefinite matrices

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 $\mathbb{R}^d$)

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

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

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

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

We saw that $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is PSD. What about $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$?

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

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

When is a diagonal matrix PSD?

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

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

If M is PSD, must cM be PSD for a constant c?

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

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

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

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 $x \in \mathbb{R}^r$. Then

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

Another useful fact: any covariance matrix is PSD.

A hierarchy of square matrices

`Square`	$M \in \mathbb{R}^{d \times d}$
$\cup$	
`Symmetric`	$M = M^T$
$\cup$	
`Positive semidefinite`	$x^T M x \geq 0$ for all $x \in \mathbb{R}^d$

Second-derivative test for convexity

A function of several variables, $F(z)$, is convex if its second-derivative matrix $H(z)$ is positive semidefinite for all z.

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

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

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

Least-squares regression

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

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