Convexity

DS-GA 1013 / MATH-GA 2824 Mathematical Tools for Data Science

Carlos Fernandez-Granda

Prerequisites

Calculus (multivariate functions)

Linear algebra (norms)

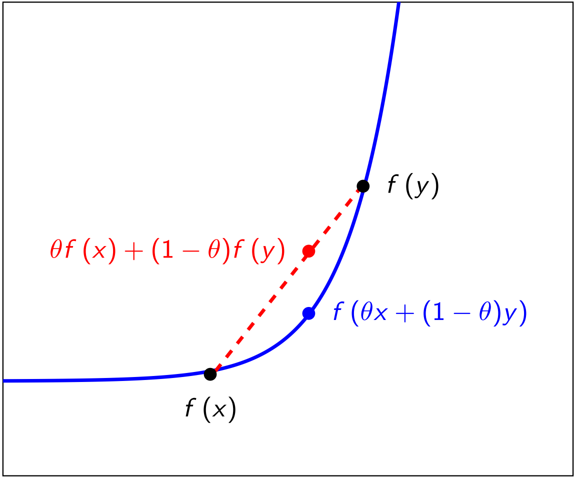
Sparse regression via the lasso

Convex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for any $x, y \in \mathbb{R}^n$ and any $\theta \in (0, 1)$

$$\theta f(x) + (1 - \theta) f(y) \geq f(\theta x + (1 - \theta) y)$$

Convex functions

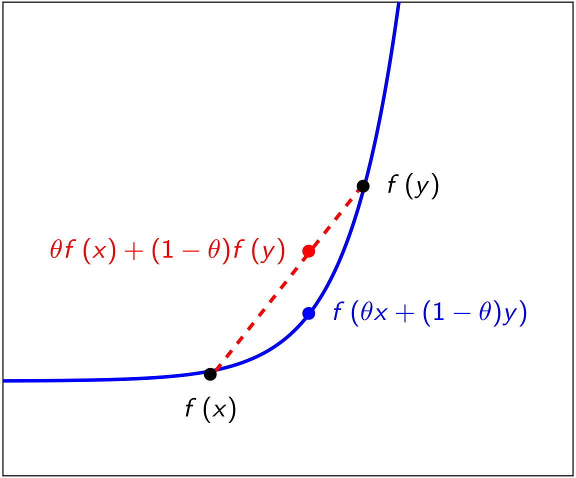


Strictly convex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if for any $x, y \in \mathbb{R}^n$ and any $\theta \in (0, 1)$

$$\theta f(x) + (1 - \theta) f(y) > f(\theta x + (1 - \theta) y)$$

Strictly convex functions



Linear functions

Linear functions are convex

$$f\left(\theta x + (1-\theta)\, y\right) = \theta f\left(x\right) + (1-\theta)\, f\left(y\right)$$

Quadratic forms

Let A be a symmetric matrix, if

$$f(x) := x^T A x \geq 0 \qquad \text{for all } x$$

then the quadratic form f is positive semidefinite

If

$$f(x) := x^T A x > 0 \qquad \text{for all } x$$

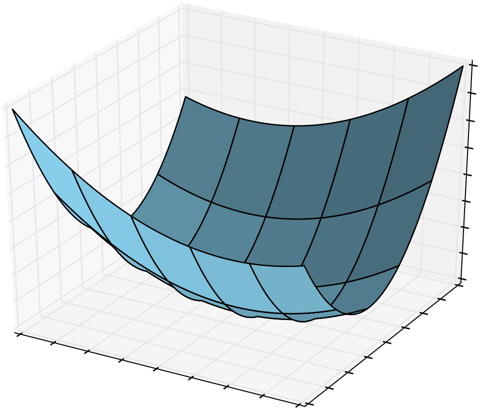
then the quadratic form f is positive definite

Positive semidefinite quadratic forms are convex

$$\begin{aligned}
&\theta f(x) + (1-\theta) f(y) - f(\theta x + (1-\theta) y) \\
&= \theta x^T A x + (1-\theta) y^T A y - (\theta x + (1-\theta) y)^T A (\theta x + (1-\theta) y) \\
&= (\theta - \theta^2) x^T A x + (1 - \theta - (1-\theta)^2) y^T A y - 2\theta(1-\theta) x^T A y \\
&= \theta(1-\theta) x^T A x + \theta(1-\theta) y^T A y - 2\theta(1-\theta) x^T A y \\
&= \theta(1-\theta)(x-y)^T A (x-y)
\end{aligned}$$

Function is convex if quadratic form is positive semidefinite, strictly convex if it is positive definite

Positive semidefinite quadratic function



Norms are convex

For any $x, y \in \mathbb{R}^n$ and any $\theta \in (0,1)$

$$\begin{aligned} ||\theta x + (1-\theta)\, y|| &\leq ||\theta x|| + ||(1-\theta)\, y|| \\ &= \theta\, ||x|| + (1-\theta)\, ||y|| \end{aligned}$$

ℓ_0 "norm" is not convex

Let $x := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $y := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, for any $\theta \in (0, 1)$

$$||\theta x + (1 - \theta) y||_0 = 2$$

$$\theta ||x||_0 + (1 - \theta) ||y||_0 = 1$$

Is the lasso cost function convex?

f strictly convex, g convex, $h := f + \lambda g$?

$$\begin{aligned} h(\theta x + (1-\theta)\,y) &= f(\theta x + (1-\theta)\,y) + \lambda g(\theta x + (1-\theta)\,y) \\ &< \theta f(x) + (1-\theta)\,f(y) + \lambda\theta g(x) + \lambda\,(1-\theta)\,g(y) \\ &= \theta h(x) + (1-\theta)\,h(y) \end{aligned}$$

Lasso cost function is convex

Sum of convex functions is convex

If at least one is strictly convex, then sum is strictly convex

Scaling by a positive factor preserves convexity

Lasso cost function is convex!

Local minima are global

Any local minimum of a convex function is also a global minimum

Proof

Let x_{loc} be a local minimum: for all $x \in \mathbb{R}^n$ such that $||x - x_{\text{loc}}||_2 \leq \gamma$

$$f(x_{\text{loc}}) \leq f(x)$$

Let x_{glob} be a global minimum

$$f(x_{\text{glob}}) < f(x_{\text{loc}})$$

Proof

Choose θ so that $x_\theta := \theta x_{\text{loc}} + (1-\theta)\, x_{\text{glob}}$ satisfies

$$||x_\theta - x_{\text{loc}}||_2 \leq \gamma$$

then

$$\begin{aligned} f\left(x_{\text{loc}}\right) &\leq f\left(x_\theta\right) \\ &= f\left(\theta x_{\text{loc}} + (1-\theta)\, x_{\text{glob}}\right) \\ &\leq \theta f\left(x_{\text{loc}}\right) + (1-\theta)\, f\left(x_{\text{glob}}\right) \quad \text{by convexity of } f \\ &< f\left(x_{\text{loc}}\right) \quad \text{because } f\left(x_{\text{glob}}\right) < f\left(x_{\text{loc}}\right) \end{aligned}$$

Strictly convex functions

Strictly convex functions have at most one global minimum

Proof: Assume two minima exist at $x \neq y$ with value $v_{\min}$

$$\begin{aligned} f(0.5x + 0.5y) &< 0.5 f(x) + 0.5 f(y) \\ &= v_{\min} \end{aligned}$$

What have we learned?

Definition of convexity

The lasso function is convex

The local minima of convex functions are global minima (and are unique for strictly convex functions)