Directional Derivatives and Optimality

David Rosenberg

New York University

January 31, 2017
Convex Sets

Definition

A set $C$ is **convex** if for any $x_1, x_2 \in C$ and any $\theta$ with $0 \leq \theta \leq 1$ we have

$$\theta x_1 + (1 - \theta) x_2 \in C.$$
Convex Sets and Functions

Definition
A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if $\text{dom } f$ is a convex set and if for all $x, y \in \text{dom } f$, and $0 \leq \theta \leq 1$, we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y).$$
Directional Derivatives and Minima
Directional Derivatives

Definition

A [one-sided] directional derivative of $f$ at $x$ in the direction $v$ is

$$f'(x; v) = \lim_{h \downarrow 0} \frac{f(x + hv) - f(x)}{h},$$

and it can be $\pm \infty$ (e.g. for discontinuous functions).

- If $f$ is convex and finite near $x$, then $f'(x; v)$ exists.
- If $f$ is differentiable then for all $v$,

$$f'(x; v) = v^T \nabla f(x).$$
Descent Directions and Optimality

Definition

\( \nu \) is a descent direction for \( f \) at \( x \) if \( f'(x; \nu) < 0 \).

- For differentiable \( f \), if \( \nabla f(x) \neq 0 \), then \( \delta x = -\nabla f(x) \) is a descent direction.
- We have a nice characterization for a minimum in terms of directional derivative:

Theorem

If \( f \) is convex and finite near \( x \), then either

- \( x \) minimizes \( f \), or
- there is a descent direction for \( f \) at \( x \).
Directional Derivatives and Minima

$\lambda_{\text{max}}$ for Lasso

- Lasso objective
  \[ J_\lambda(w) = \sum_{i=1}^{n} (w^T x_i - y_i)^2 + \lambda \|w\|_1 \]

- Is there a $\lambda_{\text{max}}$ such that $\lambda \geq \lambda_{\text{max}}$ implies $\arg\min_w J_\lambda(w) = 0$?

- Suppose yes.

- Then $w = 0$ is a minimum of $J_\lambda(w)$.

- Let’s see what that means in terms of our directional derivative characterization.
Directional Derivative for Lasso

- Consider a step direction $\nu$. For convenience, take $\nu$ s.t. $|\nu| = 1$.
- Then directional derivative at $w = 0$ in direction $\nu$ is
  \[ J'_\lambda(0; \nu) = \lim_{h \downarrow 0} \frac{J(h\nu) - J(0)}{h}. \]
- For $w = 0$ to be a minimizer, need to have $J'_\lambda(0; \nu) \geq 0$ for every direction $\nu$.
- Can find $\lambda_{\text{max}}$ by finding conditions on $\lambda$ for this to be the case.