Optimization Problems

Wednesday, April 12, 2017 6:14 PM

Empirical risk minimization

- parameters β in $\mathcal{D}$
- data $\{x_i, y_i\}$
- loss function $\ell_i(\beta) = \ell(x_i, y_i, \beta)$

$$\boxed{\hat{\beta} = \arg\min_{\beta \in \mathcal{D}} \frac{1}{n}\sum_{i=1}^{n} \ell_i(\beta) \qquad (ERM)}$$

Ex OLS: $\ell_i(\beta) = (y_i - x_i^T\beta)^2$

$\mathcal{D} = \mathbb{R}^p$

logistic regression: $y_i \in \{0,1\}$

$\ell_i(\beta) = y_i x_i^T \beta - \log(1 + e^{x_i^T\beta})$

lasso: $\ell_i(\beta) = (y_i - x_i^T\beta)^2$

$\mathcal{D} = \{ \sum_{j=1}^{p} |\beta_j| \le C \}$

such that

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 \quad \text{s.t.} \quad \sum_{j=1}^{p} |\beta_j| \le C$$

All take the form (general optimization)

$$\min_{\beta \in \mathcal{D}} R_n(\beta)$$

objective

domain

subset selection :

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \quad \text{s.t.} \quad \|\beta\|_0 \le k$$

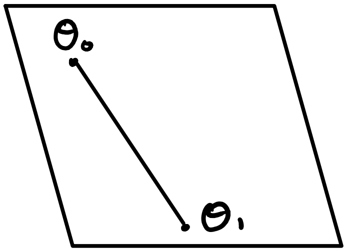
What makes this hard? non-convexity / discontinuous

Convexity

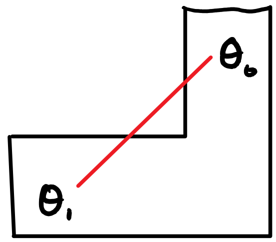
Wednesday, April 12, 2017 8:06 PM

Convex Set: $C \subseteq \mathbb{R}^p$ is convex if

$\forall\ \theta_0, \theta_1 \in C$ then $\alpha\theta_0 + (1-\alpha)\theta_1 \in C \quad \forall\, 0 \le \alpha \le 1$



Convex

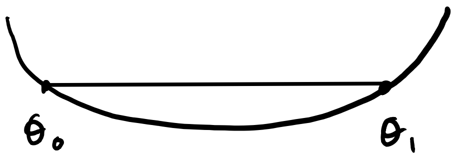


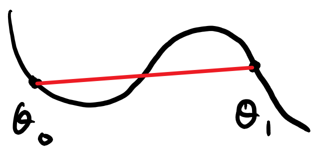
Non-convex

Convex Function: $f : \mathbb{R}^p \to \mathbb{R}$ s.t. $\mathrm{dom}(f)$ is convex

$$f(\alpha\theta_0 + (1-\alpha)\theta_1) \le \alpha f(\theta_0) + (1-\alpha) f(\theta_1)$$

for all $0 \le \alpha \le 1$, $\theta_0, \theta_1 \in \mathrm{dom}(f)$





Convex Optimization

inequality constraints

$$\min_{\beta \in D} R(\beta) \quad \text{s.t.} \quad g_i(\beta) \le 0, \ i = 1, \dots, m$$

$$h_j(\beta) = 0, \ j = 1, \dots, r$$

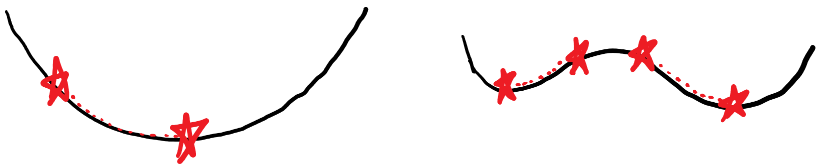
common domain

equality constraints

is convex if R, g_i are convex

& h_j is affine $(h_j(\beta) = A_j\beta + b_j)$

Property: For convex functions, local minima <u>are</u> global minima local minima <u>are</u> global minima



local minimum: β^{loc} s.t. exists $\rho > 0$ s.t.

$\forall \beta' \in D$ and $\|\beta^{loc} - \beta'\| \leq \rho$ then

$$R(\beta^{loc}) \leq R(\beta')$$

global minimum: β^{glo} s.t. $\forall \beta' \in D$

$$R(\beta^{glo}) \leq R(\beta')$$

Examples of Convex Functions

Wednesday, April 12, 2017 8:48 PM

How do we know if a function is convex?

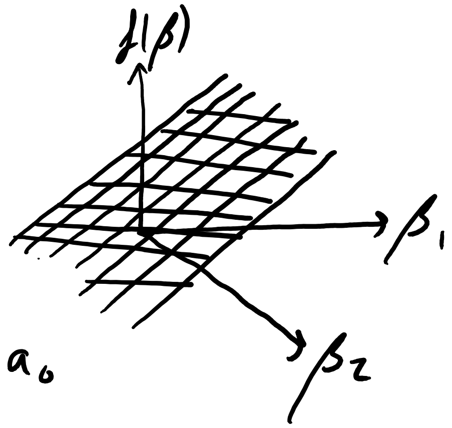
1^{st} order characterization: f is differentiable

$$f(\beta) \geq f(\beta_0) + \nabla f(\beta_0)^T(\beta - \beta_0) \quad \forall \beta, \beta_0 \in D$$

- $\nabla f(\beta_0) = 0 \Rightarrow f(\beta) \geq f(\beta_0)$! β_0 is global minimum

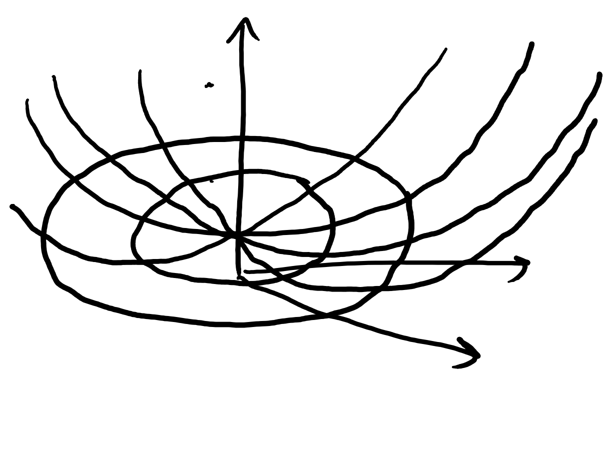
2^{nd} order characterization: f is twice differentiable

$$\nabla^2 f(\beta) \succeq 0 \quad \left(\begin{array}{c} \text{Positive Semi-definite}: A \succeq 0 \\ x^T A x \geq 0 \quad \forall x \in \mathbb{R}^p \end{array} \right)$$



Examples

- linear $f(\beta) = a^T\beta + a_0$
- quadratic $f(\beta) = \beta^T A \beta + a^T\beta + a_0$
  for PSD $A \succeq 0$



- log $f(\beta) = \log(\beta)$
- log-sum-exp

$$f(\beta) = \log\left(\sum_{i=1}^{n} \exp\left(x_i^T \beta + b_0 \right) \right)$$

- operations that preserve convexity:

  addition, partial minimization (minimize some dimensions),

  some composition $\left(\text{convex} \circ \begin{matrix} \text{convex and} \\ \text{non-decreasing} \end{matrix} \right)$

Outline of Convex Optimization Methods

Wednesday, April 12, 2017 9:19 PM

Different optimization problems (unconstrained, linear, quadratic, linear constraints, equality constraints, etc.) have different methods.

First Order Methods:

1. Gradient descent: unconstrained smooth
2. Subgradient descent: unconstrained convex
3. Proximal gradient: unconstrained convex
4. Projected gradient: constrained smooth

Second Order Methods:

1. Newton method: unconstrained smooth
2. Barrier methods: constrained smooth
3. Primal-dual interior point: linear programs
4. Quasi-Newton methods: unconstrained smooth

Many more: stochastic gradient, coordinate descent, etc.