10/8/13

DETOUR: ~~EA~~ CONSTRAINED OPTIMIZATION / LAGRANGIAN DUALITY

[NOT ~~B~~ GOING TO BE ON ANY EXAM]

1 Method of Lagrangian Multipliers for Equality Constrained Optimization

Canonical Problem: $\min_{w \in \mathbb{R}^d} f(w)$

$$h(w) = 0$$

Generalization: Can have multiple constraints

$$\left.\begin{array}{l} h_1(w) = 0 \\ h_2(w) = 0 \\ \vdots \end{array}\right\} \vec{h}(w) = \vec{0}$$

Method of Lagrangian Mults: (MLM):

$$\min_w f(w)$$

$$\left.\begin{array}{lll} h_1(w) = 0 & : & \beta_1 \\ \vdots & & \\ h_m(w) = 0 & : & \beta_n \end{array}\right\} \leftarrow \text{Lagrangian multipliers}$$

Lagrangian: $L(w, \beta_1 \ldots, \beta_n) = f(w) + \beta_1 h_1(w) + \beta_2 h_2(w) + \cdots + \beta_n h_n(w)$

Now

$$\left.\begin{array}{l} \frac{\partial L}{\partial w} = 0 \\ \frac{\partial L}{\partial \beta_1} = 0 \\ \frac{\partial L}{\partial \beta_n} = 0 \end{array}\right\} \begin{array}{l} \text{System of } n+d \text{ equations} \\ \text{in } n+d \text{ unknowns} \\ \text{Solve for } \vec{w}^*, \vec{\beta}^* \end{array}$$

(may be many)

Note 1: MLM gives critical points. Then need to check which one is global min (if any)

Note 2: $\frac{\partial L}{\partial \beta_i} = 0 \Rightarrow h_i(w) = 0 \Big\}$ Original constraint

So we are basically solving
find w s.t. $\frac{\partial L}{\partial w} = 0$

$$h_i(w) = 0 \quad \forall i$$

Note 3: For 1 constraint

$$\frac{\partial L}{\partial w} = 0 \Rightarrow \frac{\partial}{\partial w}\left[f(w) + \beta h(w) \right] = 0$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial w} = -\beta \frac{\partial h}{\partial w}}$$

This has a geometric intuition in terms of tangents. See figure on slides.

(2) MLM Example

$$\min_{w \in \mathbb{R}^d} \; w^T A w$$

$$w^T w = 1$$

$$L(w, \beta) = w^T A w + \beta(1 - w^T w)$$

$$\frac{\partial L}{\partial w} = w^T A - \beta w^T = 0$$

$$\Rightarrow \quad Aw = \beta w$$

Eigensystem

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow \quad w^T w = 1$$

$\Rightarrow$ every eigenvector, evalue is a critical point

$$f(w) = w^T A w = w^T(\beta w) = \beta$$

$\Rightarrow$ w^* = eigenvector with smallest eigenvalue.

(3) Lagrangian Duality (for general inequality/equality constrained programs)

$$\min_{w \in \mathbb{R}^d} \; f(w)$$

Lagrangian Multipliers

$\alpha \geq 0$: $\quad g(w) \leq 0$

β : $\quad h(w) = 0$

if you have a $q(w) \geq 0$ constraint, we can just define $g(w) = -q(w)$

Lagrangian $L(w, \alpha, \beta) = f(w) + \alpha g(w) + \beta h(w)$

Game-Theoretic Intepretation of Lagrangian Duality

Primal Player: Controls w
Dual Player: Controls α, β

L represents payoff from Primal player to Dual Player

Primal Player wants to $\min_{w} L(w, \alpha, \beta)$
Dual " " " $\max_{\alpha \geq 0;\, \beta} L(w, \alpha, \beta)$

2 choices

$\min_{w} \max_{\alpha \geq 0, \beta} L(w, \alpha, \beta)$

Dual-Player Goes first
Primal-Player Goes Next

$\Downarrow$

Primal Program

$\min_{w} l_p(w)$

where $l_p(w) = \max_{\alpha \geq 0, \beta} L(w, \alpha, \beta)$

$\max_{\alpha \geq 0, \beta} \min_{w} L(w, \alpha, \beta)$

Primal Player Goes first
Dual " " Next

Dual Program

$\max_{\alpha, \beta} l_D(\alpha, \beta)$

$\alpha \geq 0$

where $l_D(\alpha, \beta) = \min_{w} L(w, \alpha, \beta)$

Now $\quad l_p(w) = \max\limits_{\substack{\alpha \geq 0 \\ \beta}} L(w, \alpha, \beta)$

$$\begin{array}{ll} [\because \text{dual-player will set } \alpha = 0] & = \begin{cases} f(w) & \text{if } g(w) \leq 0,\ h(w) = 0 \\ \infty & \text{else} \end{cases} \\ \left[\begin{array}{l}\because \text{dual-player will set } \alpha = +\infty \\ \quad \text{and/or } \beta = \pm\infty\end{array}\right] & \end{array}$$

So Primal - Program:

$$\min_w l_p(w) \equiv \left.\begin{array}{l} \min\limits_w f(w) \\ g(w) \leq 0 \\ h(w) = 0 \end{array}\right\}$$

Basically, by going first the dual player extracts $+\infty$ penalty from Primal player for violating constraints. Thus Primal player always satisfies constraints.

Dual Program:

$$\max_{\substack{\alpha \geq 0 \\ \beta}} l_D(\alpha, \beta)$$

$$l_D(\alpha, \beta) = \min_w \left[f(w) + \alpha g(w) + \beta h(w) \right]$$

Because Primal player goes first, he can violate constraints as long as the penalty for violation (α, β) is NOT ∞. Dual player may not set ∞ penalties because he may be able to maximize payoff at some NON-INF values.

④ Relationship between primal & dual values

Let $\underbrace{p^*}_{\text{primal OPT value}} = \min_{\substack{g(w) \le 0 \\ h(w)=0}} f(w)$ $\qquad$ $\underbrace{d^*}_{\text{Dual OPT value}} = \max_{\substack{\alpha \ge 0 \\ \beta}} \ell_D(\alpha, \beta)$

$p^* \geq d^*$ [Weak Duality]

Always true, without any assumptions on f, g, h, etc.

Why?

$p^* = \min \max L$

$d^* = \max \min L$

min max Something $\geq$ max min Something

Example

$$\begin{array}{c|cc|c} w \backslash \alpha & & & \max_\alpha \\ \hline & 0 & 1 & 1 \\ & 1 & 0 & 1 \\ \hline \min_w & 0 & 0 & \end{array}$$

In general

$f(w)$ } Any feasible w can only have $f(w) \geq p^*$ because p^* is min

Weak Duality { p^*, d^*

d^*

$l_D(\alpha, \beta)$ } Any ~~feasible~~ α, β can only have $l_D(\alpha, \beta) \leq d^*$ because d^* is max.

So $l_D(\alpha, \beta)$ for ANY $\alpha \geq 0, \beta$ provides a lower-bound on p^* & $f(w)$ for ANY w.

④ KKT Conditions & Strong Duality

Let $w^* = \underset{\substack{g(w) \leq 0 \\ h(w) = 0}}{\operatorname{argmin}} f(w)$ } Technical point: argmin may not exist [Ignore in ML class]

Let $\alpha^*, \beta^* = \underset{\alpha \geq 0}{\operatorname{argmax}} \; l_D(\alpha, \beta)$

Karush-Kuhn-Tucker (KKT) Conditions

$p^* = d^*$

w^* is P-OPT

α^*, d^* is D-OPT

STRONG DUALITY

$\xrightarrow{\text{necessary}}$

$\xleftarrow{\text{sufficient IF}}$ f, g convex; h affine & $\exists w$ s.t. $g(w) < 0$ [SLATOR'S CONDITION]

$\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial w} = 0$

$\alpha^* g(w^*) = 0$ [Complimentary Slackness]

$\alpha^* \geq 0$ Dual Feasibility

$g(w^*) \leq 0$, $h(w^*) = 0$ } Primal Feasibility