①

Constrained optimization:

$$\min f(x) \quad st \quad c_i(x) = 0$$
$$g_i(x) \geqslant 0$$

Lagrangian

$$\mathcal{L}(x, \underline{\lambda}) = f(x) - \lambda^{(e)T} \underline{c} - \underline{\lambda}^{(i)T} \underline{g}$$

(here λ is a vector of constraints whose elements csp to ~~to~~ ineq ($\lambda^{(i)}$) or eq constraints ($\lambda^{(e)}$))

Necessary conditions (KKT conds)

$$\nabla_x \mathcal{L} = 0 \qquad g_i(x) \geqslant 0$$

$$c_i(x) = 0 \qquad \lambda^{(i)} \geqslant 0$$

$$\lambda_i^{(e)} c_i = 0 \qquad \lambda_i^{(i)} g_i = 0$$

Duality

example:

$$\min \frac{x^T x}{2} \quad \text{st} \quad Ax = b$$

$$\mathcal{L}(x,\lambda) = \frac{x^T x}{2} - \lambda^T (Ax - b)$$

from first condition:

$$x^T - \lambda^T A = 0$$

i.e. $$x = A^T \lambda$$

substitute

$$\frac{\lambda^T A A^T \lambda}{2} - \lambda^T (A A^T \lambda - b)$$

$\rightarrow \lambda \rightarrow x$

Knowledge of LM values is powerful!

- Assume inequality constraints only.

$$\min_x f(x) \quad \text{st} \quad g_i(x) \geq 0$$

- Assume $-g_i$ is convex

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T g(x)$$

Define dual objective fn to be

$$q(\lambda) = \inf_x \mathcal{L}(x, \lambda).$$

on domain such that $q(\lambda) > -\infty$

dual problem:

$$\max_\lambda q(\lambda), \quad \lambda \geq 0$$

<u>Thm</u>: q is concave, domain is convex
(straightforward)

<u>Thm</u>: for feasible x, any λ

$$q(\lambda) \leq f(x)$$

(straightforward)

<u>Thm</u>: suppose x is soln of primal, f and $-g_i$ are convex; then λ such that (x, λ) satisfies KKT is a soln of dual

~~Thm: with~~ other way round requires stronger technical conds

<u>Thm</u>: value of dual $\leq$ value of primal.

Duals:

$$q(\lambda) = \inf_{x \in \text{feasible}} \mathcal{L}(x, \lambda).$$

1) for feasible x, any ~~feast~~ $\lambda \geq 0$

$$q(\lambda) \leq f(x)$$

$\rightarrow$ for feasible x, $g(x) \geq 0$

$\rightarrow$ $\lambda \geq 0$

so $$q(\lambda) = \inf_{x \in \text{feasible}} \left[f(x) - h(\lambda, x) \right]$$

$h(\lambda, x) \geq 0$

Dual problem:

$$\max_{\lambda \geq 0} \; q(\lambda)$$

The value of this problem $\leq$ value of primal

because

$q(\lambda) \leq f(x)$ for feasible x

Dual of a linear program

write as

$$\min \; c^T x$$
$$\text{st} \quad Ax \geq b$$

Lagrangian:

$$\mathcal{L}(x, \lambda) = c^T x - \lambda^T (Ax - b)$$

Dual:

$$\inf_x \mathcal{L}(x, \lambda)$$

But this is $-\infty$ unless

$$A^T \lambda = c$$

So we have

$$\max \quad b^T \lambda$$

$$\text{s.t.} \quad A^T \lambda = c$$

$$\lambda \geq 0$$

(Notice this has a <u>slightly</u> different form than primal; if we write primal as

$$\min \quad c^T \lambda$$

$$\text{st} \quad Ax \geq b$$

$$x \geq 0$$

we get a dual of similar form.

Equality constraints

assume we have

$$\min f(x)$$
$$\text{st} \quad g(x) = 0$$

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T g(x)$$

among the KKT, we find.

$$\nabla f - \lambda^T J_g = 0 \quad \leftarrow \quad \nabla_x \mathcal{L} = 0$$

AND $$g(x) = 0 \quad \longleftarrow \quad \nabla_\lambda \mathcal{L} = 0$$

So this is a stationary point of $\mathcal{L}$ in (x, λ).

(

back to simple model prob

$$\min \frac{x^T x}{2}$$

$$\text{st} \quad Ax = b$$

short + fat.

$$\mathcal{L}(x, \lambda) = \frac{x^T x}{2} - \lambda^T (Ax - b)$$

$$x = A^T \lambda \qquad Ax = b$$

$$\begin{pmatrix} Id & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}$$

1) we could solve this as is.

2) we could eliminate x

$$A(A^T \lambda) = b \longrightarrow \lambda \longrightarrow x$$

smaller than x

Common application: in important cases, one may be able to write the dual directly.

SVM

$$\left.\begin{array}{l} \min \quad \frac{w'w}{2} \\ \text{st } y_i(w'x_i + b) \geq 1 \end{array}\right] \text{Primal form, Separable}$$

$$\mathcal{L}_P(w,\lambda) = \frac{w'w}{2} - \sum_i \lambda_i \{[y_i(w'x_i + b)] - 1\}$$

$$\nabla_w \mathcal{L} = 0 = w - \sum_i \lambda_i \{[y_i x_i]\}$$

$$\nabla_b \mathcal{L} = 0 = -\sum_i \lambda_i y_i$$

Subst

$$\mathcal{L}_D = \sum_i \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j \left[y_i y_j (x_i^T x_j) \right]$$

Notice constraints

$$\sum_i \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

and we must max this in λ

If there is an fp for primal, the max is soln to primal

i.e. Value (Dual) = Value (Primal)

What if data is not separable?

$$\left.\begin{aligned} \min\ & \frac{w'w}{2} + C\sum_i \xi_i \\ \text{st}\ & y_i(w'x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned}\right] \text{ Primal prob}$$

ξ_i are <u>slack variables</u>

$$\mathcal{L}_p = \frac{w'w}{2} + C\sum_i \xi_i - \sum_i \lambda_i [y_i(w'x_i + b) - 1 + \xi_i] - \sum_i \mu_i \xi_i$$

$$\nabla_w \mathcal{L}_p = w - \sum_i \lambda_i y_i x_i = 0$$

$$\nabla_b \mathcal{L}_p = 0 = -\sum_i \lambda_i y_i$$

$$\nabla_{\xi_i} \mathcal{L}_p = C - \lambda_i - \mu_i = 0 \Big] \rightarrow \text{this gets rid of } \xi_i$$

So we have

$$\mathcal{L}_D = \sum_i \lambda_i - \frac{1}{2} \sum_{ij} y_i y_j \lambda_i \lambda_j x_i' x_j$$

subject to

$$\sum_i \lambda_i y_i = 0$$

$$0 \leq \lambda_i \leq C$$

Notice that ξ_i can be interpreted as a <u>loss</u>

$$\text{hinge loss}(y, y_p) = \max(0, 1 - y_i y_p)$$

Methods:

Quadratic penalty method

(assume equalities)

$$\min_x f(x) + \frac{\mu}{2} \sum_i c_i^2(x) = Q_K(x)$$

and drive $\mu \to \infty$, resolve

Notice at soln

$$\nabla_x Q_K \simeq 0 = \nabla f + \sum_i \left(\mu_K c_i(x_*)\right) \nabla c_i(x)$$

By inspection, this would match $\nabla_x \mathcal{L} \simeq 0$, if

$$-\mu_K c_i = \lambda_i^*$$

Which suggests that at conv $c_i = \frac{-\lambda_i^*}{\mu_K}$

This looks OK, because $\mu_K \to \infty$, but not exact. Also, $\mu_K \to \infty$ creates major probs w/ Hessian

Augmented Lagrangian method

Consider

$$\mathcal{L}_A(x, \lambda; \mu) = f - \sum_i \lambda_i c_i + \frac{\mu}{2} \sum_i c_i^2$$

- have an est of λ^K, μ_K, get x^*

- at x^* $\quad \nabla_x \mathcal{L}_A = 0 = \nabla f - \sum_i (\lambda_i^K - \mu_K c_i) \nabla c_i$

- This suggests $\quad \lambda_i^* \approx (\lambda_i^K - \mu_K c_i)$

  and $\quad c_i \approx -\frac{1}{\mu_K} \cdot [\lambda_i^* - \lambda_i^K]$

which suggests moving $\lambda_i \to \lambda_i^*$

But we have a good est:

$$\lambda_i^* \approx (\lambda_i^k - \mu_k c_i)$$

so update ests, go again.

1) Method converges w/o increasing μ_k indefinitely