Constrained optimization:

$$\min f(x) \quad \text{st} \quad c_i(x) = 0$$
$$g_i(x) \geq 0$$

Lagrangian

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

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

Necessary conditions (KKT conds)

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

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

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

- Assume inequality constraints only.

<u>Duality</u>

$\min f(x)$ st $g_i(x) \geq 0$

<u>example</u>:

- Assume $\min \frac{x^T x}{2}$ is st $\underline{\text{conv}}$ $\cancel{Ax} = b$

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

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

define dual objective fn to be

from first condition:

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

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

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

substitute

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

<u>dual problem</u>:

$\lambda \rightarrow x$

$\max q(\lambda)$ $\lambda \geq 0$

Knowledge of λ w.r.t values is powerful!

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

<u>Thm</u>: for feasible x, any λ

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

(straight forward)

<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

~~<u>Thm</u>: with~~ Other way round requires stronger technical conds

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

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

SVM

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

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

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

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

Subst :

~~Nun~~

P

<u>subst</u>

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

Notice constraints

$$\sum_i \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

and we must <u>max</u> this in λ

<u>If</u> there is an fp for primal, the max is soln to primal

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

What if data is not separable? (7)

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

ξ_i are $\underline{\text{slack variables}}$

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

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

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

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

Next we have

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

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

$$\nabla_x Q_k \approx 0 = \nabla f + \sum_i \left(\mu_k c_i(x_k) \right) \nabla c_i(x)$$

By inspection, this would match $\nabla_x \mathcal{L} = 0$, 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 $\quad \lambda_i^* \approx (\lambda_i^k - \mu_k c_i)$

$$\text{and} \quad c_i \approx -\frac{1}{\mu_k} \cdot [\lambda_i^* - \lambda_i^k]$$

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