

today: Lagrangian duality
dual of SVM struct

review of Lagrangian duality

primal problem:
 $p^* = \inf_x f(x)$
s.t. $h_i(x) \leq 0 \forall i \in I$

dual problem
 $d^* := \sup_{\alpha \geq 0} d(\alpha)$
Lagrangian dual fct.

saddle point perspective: $\inf_x f(x)$
s.t. $h_i(x) \leq 0 \forall i \in I$

(see Lecture 15 of Feb 2016)

$p^* = \inf_x \sup_{\alpha \geq 0} g(x, \alpha)$
"weak duality" \rightarrow always true
 $\geq \sup_{\alpha \geq 0} \inf_x g(x, \alpha) \triangleq d^*$
 $\triangleq d(\alpha)$
Lagrangian dual fct.

KKT conditions \rightarrow necessary conditions for x^* & α^* to be primal and dual optimum
when f & h_i are differentiable functions

- x^* primal feasible (ie. $h_i(x) \leq 0 \forall i \in I$)
- α^* dual feasible ($\alpha^* \geq 0$)

Fenchel conjugate of $f \rightarrow f^*(y) \triangleq \sup_x \langle y, x \rangle - f(x)$
 \downarrow
 $x^* \text{ s.t. } \nabla f(x^*) = y$

aside Fenchel duality:
 $p^* = \inf_x f(x) + g(Ax)$

$d^* = \sup_y f^*(A^T y) - g^*(y)$

$g(x, \alpha)$ "Lagrangian fct."
 $= \inf_x \left[\sup_{\alpha \geq 0} f(x) + \sum_i \alpha_i h_i(x) \right]$

because $g(x) \triangleq \sup_{\alpha \geq 0} \sum_i \alpha_i h_i(x) = \begin{cases} +\infty & \text{if } \exists i \text{ s.t. } h_i(x) > 0 \\ 0 & \text{o.w.} \end{cases}$
i.e. if x is feasible

we have $g(x) = \delta_C(x)$

where C is feasible set

$$\left\{ \begin{array}{l} \bullet x^* \text{ primal feasible (i.e. } h_i(x) \leq 0 \forall i \in I) \\ \bullet \alpha^* \text{ dual feasible } (\alpha^* \geq 0) \\ \bullet \nabla f(x^*) + \sum_i \alpha_i^* \nabla h_i(x^*) = 0 \quad (\text{i.e. } (x^*, \alpha^*) \text{ stationary point of } \mathcal{L}(x, \alpha) \text{ i.e. } \nabla_x \mathcal{L}(x^*, \alpha^*) = 0) \\ \bullet \alpha_i^* h_i(x^*) = 0 \forall i \quad \text{complementary slackness} \end{array} \right.$$

if primal problem is convex (i.e. f & h_i are convex fct.)

(constraint qualification condition)

and " Slater's condition" \Rightarrow

strong duality i.e. $p^* = d^*$

and KKT conditions are sufficient (i.e. here can get α^* , then x^* through KKT and then know that x^* solves primal)

$$\boxed{\begin{array}{l} \exists x \in \text{relint}(\text{dom}(f)) \\ \text{s.t. } h_i(x) < 0 \forall i \\ \text{where } h_i \text{ is not affine} \end{array}}$$

by weak duality, $f(x) \geq d(\alpha) \forall x \text{ \& } \alpha \text{ feasible}$

$d(\alpha)$ always concave in α even if f is non-convex

(inf of concave functions is concave)

$d(\alpha)$ is obtained through unconstrained optimization

counter example for strong duality (see Boyd's book exercise 5.21)

$$\begin{array}{ll} \min f(x, y) & \text{where } f(x, y) = \begin{cases} x^2/y & y > 0 \\ +\infty & \text{o.w.} \end{cases} \\ \text{s.t. } x^2/y \leq 0 \end{array}$$

no unique "Lagrangian dual problem"
 \rightarrow depends on how write the primal problem

Structured SVM dual problem

Conv. primal: $\min \|x\|_w^2 \text{ s.t. } \sum_{i=1}^N \xi_i$

unconst. primal: $\min \|x\|_w^2 + \lambda \sum_{i=1}^N f_i(w)$

Conv. primal: $\min_w \frac{\lambda \|w\|^2}{2} + \sum_{i=1}^n \xi_i$
 $\xi_i \geq H_i(\tilde{y}; w) \quad \forall \tilde{y} \in \mathcal{Y}_i$

Sagravian
duality

$\ell_i(\tilde{y}) - w^T \psi_i(\tilde{y})$

unconst. primal: $\min_w \frac{\lambda \|w\|^2}{2} + \sum_{i=1}^n H_i(w)$

Fenchel duality...

$L(w, \xi; \alpha) = \frac{\lambda \|w\|^2}{2} + \sum_{i=1}^n \xi_i + \underbrace{\frac{1}{n} \sum_{i=1}^n \alpha_i(\tilde{y})}_{\text{high } w, \xi \leq 0 \text{ constraint}} \left[\ell_i(\tilde{y}) - w^T \psi_i(\tilde{y}) - \xi_i \right]$

$\frac{\partial L(w, \xi; \alpha)}{\partial \xi_i} = 0 \Rightarrow \boxed{\sum_{\tilde{y} \in \mathcal{Y}_i} \alpha_i(\tilde{y}) = 1} \quad \text{i.e. } \alpha_i \in \Delta_{|\mathcal{Y}_i|}$

$\nabla_w L = 0 \Rightarrow \lambda w - \frac{1}{n} \sum_{i=1}^n \alpha_i(\tilde{y}) \psi_i(\tilde{y}) = 0$

$w(\alpha) \triangleq \frac{1}{\lambda n} \sum_{i=1}^n \sum_{\tilde{y} \in \mathcal{Y}_i} \alpha_i(\tilde{y}) \psi_i(\tilde{y})$

KKT condition

$d(\alpha) = \frac{\lambda \|w(\alpha)\|^2}{2} + \underbrace{\frac{1}{n} \sum_{i=1}^n \alpha_i(\tilde{y}) [\ell_i(\tilde{y}) - w(\alpha)^T \psi_i(\tilde{y})]}_{\langle b, \alpha \rangle}$

notation: $\alpha \in \Delta_{|\mathcal{Y}_1|} \times \Delta_{|\mathcal{Y}_2|} \times \dots \times \Delta_{|\mathcal{Y}_n|}$

$b = (b_i)_{i=1}^n \quad b_i(\tilde{y}) \triangleq \int \ell_i(\tilde{y})$

$d(\alpha) = \langle b, \alpha \rangle + \frac{\lambda \|w(\alpha)\|^2}{2} - w(\alpha)^T \left(\underbrace{\frac{1}{n} \sum_{i=1}^n \alpha_i(\tilde{y}) \psi_i(\tilde{y})}_{= \lambda w(\alpha)} \right)$

$\dim(\alpha) \triangleq m$
 $b \in \mathbb{R}^m$

$A \in \mathbb{R}^{d \times m}$

$A = \left[\dots \underbrace{\psi_i(\tilde{y})}_{\lambda n} \dots \right]$

$d(\alpha) = \langle b, \alpha \rangle - \frac{\lambda \|A\alpha\|^2}{2}$

dual problem: $\max -\frac{\lambda \|A\alpha\|^2}{2} + \langle b, \alpha \rangle$

dual problem:
$$\max_{\substack{\alpha_i \in \Delta(S_i) \\ 1 \leq i \leq n}} -\lambda \frac{\|A\alpha\|^2}{2} + \langle b, \alpha \rangle$$

(exponentially many dual variables?)

$$A = L \begin{bmatrix} \varphi_1(y) & \dots & \varphi_n(y) \end{bmatrix}$$

$$w(q) = A\alpha$$

Complementary slackness: $\alpha_i^*(\tilde{y}) > 0 \Rightarrow h_{i\tilde{y}}(w^*, \alpha_i^*) = 0$

i.e. $\alpha_i^* = \ell_i(\tilde{y}) - w^{*T} \varphi_i(\tilde{y}) = H_i(\tilde{y}; w^*)$

by primal feasibility, $\alpha_i^* \geq H_i(\tilde{y}, w^*) \quad \forall \tilde{y} \in S_i$

thus
$$\alpha_i^* = \max_{\tilde{y} \in S_i} H_i(\tilde{y}, w^*)$$

→ because $\varphi(y^{(i)}) = 0$

Support vectors are when $\alpha_i^*(\tilde{y}) > 0$ (and $\tilde{y} \neq y^{(i)}$)

[necessary condition is $\tilde{y} \in \arg\max_{\tilde{y} \in S_i} H_i(\tilde{y}; w^*)$]

→ $\varphi_i(\tilde{y})$

(*) gap = α^* selection
primal objective

primal-dual gap:

$$\text{gap}(\alpha) = p(w(\alpha), \alpha(\alpha)) - d(\alpha)$$

$$= \underbrace{p(w(\alpha), \alpha(\alpha)) - p(w^*, \alpha^*)}_{\text{primal suboptimality}} + \underbrace{p(w^*, \alpha^*) - d(\alpha)}_{\text{dual suboptimality}}$$

$$= \lambda \frac{\|w\|^2}{2} + \sum_{i=1}^n \alpha_i H_i(w) + \lambda \frac{\|w\|^2}{2} - \sum_{i=1}^n \alpha_i \ell_i(y)$$

$$= \frac{1}{n} \sum_{i=1}^n [H_i(\omega) - \sum_y \alpha_i(y) f_i(y)] + \lambda \langle \omega, \underbrace{\omega(\alpha)}_{} \rangle$$

$$\frac{1}{n} \sum_{i=1}^n \alpha_i(y) f_i(y)$$

$$\boxed{\text{gap}(\alpha) = \frac{1}{n} \sum_{i=1}^n \left[H_i(\omega) - \sum_y \alpha_i(y) \underbrace{f_i(y; \omega)}_{\substack{\text{max}_{\omega} f_i(y; \omega)}} \right]}$$

block α_i

→ use to get
bound on dual suboptimality of α
and primal ω of $\omega(\alpha)$

→ motivate block-gap sampling schemes
for block-coordinate ascent method