Lecture 14 - dual SVM

Tuesday, March 9, 2021 14:26

<u>today</u>
- IPM
- duality
- SVM dual

<u>interior point method:</u>

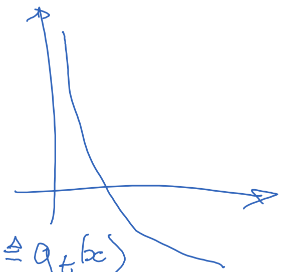
$$\min f(x) \quad \text{s.t. } Ax = b,\ x \geq 0$$

barrier method →

$t \geq 0$

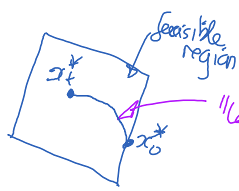
barrier fct.

$$\min f(x) + t\underbrace{\left(-\mathbb{1}^T \log x\right)}_{\sum_i \log x_i} \triangleq g_t(x) \quad \text{s.t. } Ax = b$$



$-\log x$

$$\text{dom}(g_t(x)) = \{x : x > 0\}$$



$$x_t^* = \arg\min g_t(x) \quad \text{s.t. } Ax = b$$

"central path"

Interior point method: run Newton's method <u>with warm start</u> on $g_t(x)$ and decrease t with a geometric schedule

there is no condition # of g_t appearing

"self-concordant analysis"

<u>Convergence</u>: Newton's method take $\log(\log(\frac{1}{\varepsilon})) + \text{cst}\cdot(g_t(x_0) - g_t(x_t^*))$ iterations to reach ε-accuracy

IPM: total # of Newton iterations is $\log(\frac{1}{\varepsilon}) \cdot$ some constants

↳ dimensionality is appearing; but <u>not</u> the condition # of f (vs. gradient methods)

Review of Lagrangian duality

<u>primal problem</u>

$$p^* \triangleq \inf_x f(x) \quad \text{s.t. } h_i(x) \leq 0 \ \forall i \in I$$

<u>dual problem</u>

$$d^* = \sup_{\alpha \geq 0} d(\alpha)$$

Lagrangian dual fct.

$$d(\alpha) = \inf f(x) + \sum \alpha_i h_i(x)$$

Fenchel duality

Fenchel conjugate of f

$$f^*(y) \triangleq \sup_x \langle y, x\rangle - f(x)$$

x^* is s.t. $\nabla f(x^*) = y$

<u>Fenchel duality:</u>

$$p^* = \inf_{x \in \mathbb{R}^d} f(x) + g(Ax) \qquad A: \mathbb{R}^d \to \mathbb{R}^p$$

e.g. $\frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_i H_i(w)$

$$d(\alpha) = \inf_x f(x) + \sum_i \alpha_i h_i(x)$$

$x \in \mathbb{R}^d$ e.g. $\to \frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_i H_i(w)$

$$d^* = \sup_{y \in \mathbb{R}^p} -f^*(A^T y) - g^*(y)$$

$\max_{y} g_i(y) - \langle w, \varphi_i(y) \rangle$

note: <u>not unique?</u>

saddle point perspective of Lag. duality

$$\inf_x f(x) \text{ s.t. } h_i(x) \leq 0 \ \forall i \in I = \inf_x \Big[\sup_{\alpha \geq 0} \underbrace{f(x) + \sum_i \alpha_i h_i(x)}_{\mathcal{L}(x,\alpha)} \Big]$$

Lagrange multiplier (α_i)

"Lagrangian fct."

because $g(x) \triangleq \sup_{\alpha \geq 0} \sum_i \alpha_i h_i(x) = \begin{cases} +\infty & \text{if } x \text{ is not feasible, i.e. } \exists i \text{ s.t. } h_i(x) > 0 \\ 0 & \text{o.w.} \end{cases}$

$= \delta_C(x)$ where $C \triangleq \{x : h_i(x) \leq 0 \ \forall i\}$

$$p^* = \inf_x f(x) + \delta_C(x)$$
$$= \inf_x \sup_{\alpha \geq 0} \mathcal{L}(x,\alpha)$$
$$\geq \sup_{\alpha \geq 0} \underbrace{\inf_x \mathcal{L}(x,\alpha)}_{\triangleq d(\alpha)} \triangleq d^*$$

"weak duality"

Lagrangian dual fct.

$$\inf_x \Big(\sup_{\alpha \geq 0} \mathcal{L}(x,\alpha)\Big) \geq \sup_{\alpha \geq 0} \inf_{x'} \mathcal{L}(x',\alpha)$$

strong duality $p^* = d^*$

<u>KKT conditions:</u> $\to$ necessary conditions for x^* & α^* to be primal & dual optimal when f & h_i are differentiable fcts.

KKT conditions

- x^* is primal feasible (i.e. $h_i(x^*) \leq 0 \ \forall i$)
- α^* is dual feasible ($\alpha^* \geq 0$)
- $\nabla f(x^*) + \sum_i \alpha_i^* \nabla h_i(x^*) = 0$ [i.e. (x^*, α^*) stationary pt. of $\mathcal{L}(x,\alpha)$ with respect to x]
- $\alpha_i^* h_i(x^*) = 0 \ \forall i$ "complementary slackness"

$\hookrightarrow \alpha_i^* > 0$ <u>only</u> on <u>active</u> constraints $h_i(x) = 0$

$*$ if primal problem is convex (i.e. f & h_i's are convex) <u>and</u> "Slater's condition" $\Rightarrow$ <u>strong duality</u> i.e. $p^* = d^*$

$\boxed{\exists x \text{ rel interior}(\text{dom} f)}$

<u>and</u> KKT conditions are <u>sufficient</u>

(e.g. thus suppose get α^* by solving dual,)

$$\boxed{\begin{array}{c}\exists x \text{ rel interior (dom } f) \\ \text{s.t. } h_i(x) < 0 \ \forall i \\ \text{when } h_i \text{ is not affine}\end{array}}$$

(e.g. thus suppose get α^* by solving dual, then get x^* through KKT conditions $\Rightarrow x^*$ solve primal)

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

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

$$\text{s.t. } \frac{x^2}{y} \leq 0$$

violates Slater's condition and no strong duality

<u>other comments</u>:

- $f(x) \geq d(\alpha) \quad \forall x \,\&\, \alpha$ feasible
- $d(\alpha)$ is always concave in α
- $d(\alpha)$ is obtained through <u>unconstrained</u> opt.
- no unique "Lagrangian dual problem"

e.g. $h_i(x) \leq 0 \Leftrightarrow a h_i(x) \leq 0$ for some $a > 0$

15h40

SVM struct dual problem

consider primal: $\min_{w,\xi} \frac{\lambda \|w\|^2}{2} + \frac{1}{n}\sum_{i=1}^n \xi_i$

$$\text{s.t. } \frac{1}{n}\xi_i \geq \frac{1}{n} H_i(\tilde{y}; w) \quad \forall \tilde{y} \in \mathcal{Y}_i$$

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

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

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

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

$$\boxed{w(\alpha) \triangleq \frac{1}{\lambda n} \sum \alpha_i(\tilde{y})\, \psi_i(\tilde{y})} \quad \text{KKT condition}$$

$$\boxed{w(\alpha) \triangleq \frac{1}{\lambda n} \sum_{i,\tilde{y}} \alpha_i(\tilde{y})\, \psi_i(\tilde{y})}$$ KKT condition

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

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

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

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

$b \in \mathbb{R}^m \leftarrow$ exponentially big!

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

$\boxed{w(\alpha) = A\alpha}$

$$A \triangleq \begin{array}{c} \scriptstyle i,\tilde{y} \text{ column} \\ \left[\cdots \frac{\psi_i(\tilde{y})}{\lambda n} \cdots \right] \end{array}$$

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

dual problem: $$\boxed{\max_{\substack{\alpha_i \in \Delta_{|\mathcal{Y}_i|} \\ 1 \le i \le n}} -\frac{\lambda}{2} \|\underbrace{A\alpha}_{w(\alpha)}\|^2 + \langle b, \alpha \rangle}$$

(exponentially many dual variables)

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

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

define $\xi_i(\alpha) = \max_{\tilde{y}} H_i(\tilde{y}; w(\alpha))$

by primal feasi., $\xi_i^* \ge H_i(\tilde{y}'; w^*) \ \forall \tilde{y}' \in \mathcal{Y}_i$

thus $\boxed{\xi_i^* = \max_{\tilde{y} \in \mathcal{Y}_i} H_i(\tilde{y}; w^*) = H_i(w^*)}$

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

[necessary condition $\tilde{y} \in \operatorname{argmax}_{\tilde{y} \in \mathcal{Y}_i} H_i(\tilde{y}; w^*)$]

$\circledast$ sparse solution α^*