Lecture 14 - dual SVM

Thursday, March 9, 2023 1:32 PM

today:
- IPM
- duality
- SVM dual

interior point method:

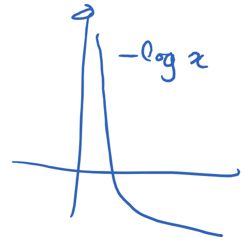
$$\min f(x) \quad \text{s.t. } Ax=b, \; x \geq 0$$

barrier method →

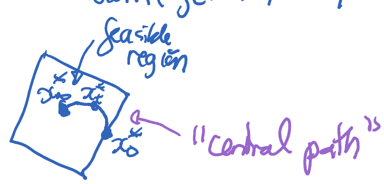
$t \geq 0$ barrier fct.

$$\min f(x) + t\,(\underbrace{-\mathbb{1}^T \log x}_{\sum_j \log x_j}) \triangleq g_t(x)$$

$$\text{s.t. } Ax=b$$



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



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

interior point method: run Newton's method with warm start on $g_t(x)$ and decrease t with a geometric schedule

Convergence: Newton's method take $\underbrace{\log(\log(\tfrac{1}{\varepsilon})) + \text{cst.}\,(g_t(x_0) - g_t(x_t^*))}$ iterations to reach ε-accuracy

there is no condition #
"self-concordant analysis"

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

↳ dimensionality appears here

but not condition # of f

(vs. gradient methods)

Review of Lagrangian duality

primal problem

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

dual problem

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

↑ Lagrangian dual fct.

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

Fenchel duality

Fenchel conjugate of f

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

↑ $x^*(y)$ s.t. $\nabla f(x^*) = y$

Fenchel duality

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

$A: \mathbb{R}^d \to \mathbb{R}^q$

e.g. $\lambda \|x\|^2 + \frac{1}{n}\sum_{i=1}^{n} \ell_i(x)$

$\alpha \geq 0$

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

$A: \mathbb{R}^d \to \mathbb{R}^q$

$p^* = \inf_{x \in \mathbb{R}^d} f(x) + g(Ax)$ → e.g. $\frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} H_i(w)$

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

<u>not unique?</u>

saddle pt. perspective of Lag. duality

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

"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 $C \triangleq \{x : h_i(x) \leq 0 \;\forall i\}$

$$p^* = \inf_x f(x) + \delta_C(x)$$

$$= \inf_x \Big[\sup_{\alpha \geq 0} \mathcal{L}(x,\alpha) \Big]$$

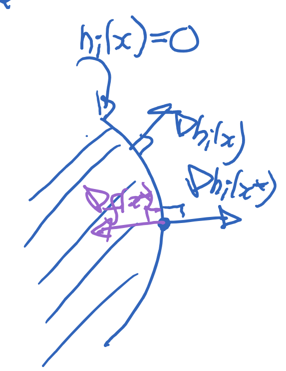
$$\geq \sup_{\alpha \geq 0} \Big[\underbrace{\inf_x \mathcal{L}(x,\alpha)}_{\triangleq d(\alpha)} \Big] \triangleq d^*$$

"weak duality" always true

Lagrangian dual fct.

strong duality $p^* = d^*$

<u>KKT conditions</u>: → necessary conditions for x^* & α^* to be primal & dual optimal when f & h_i are diff. 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"
  ↳ $\alpha_i^* > 0$ <u>only</u> on <u>active constraints</u> $h_i(x^*) = 0$

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

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

and KKT conditions are sufficient

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

(e.g. thus supposed α^* by solving dual; then get x^* through KKT condition $\Rightarrow x^*$ solve primal)

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

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

violates Slater's condition and no strong duality

other comments

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

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

14h51

SVM struct dual problem

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

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

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

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

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

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

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

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

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

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

$b = (b_i)_{i=1}^n \quad b_i(\tilde{y}) \triangleq \frac{1}{n}\ell_i(\tilde{y}) \qquad b \in \mathbb{R}^m$ ← exponentially big

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

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

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

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

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

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

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 $\tilde{y} \in \arg\max H_i(\tilde{y}; w^*)$]

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

(*) sparse solution α^*