Lecture 14 - dual SVM

Thursday, February 21, 2019 13:32

today:
- IPM
- SVM dual

interior point method:

$$\min \ f(x) \quad \text{s.t.} \ Ax=b,\ x\geq 0 \quad \xrightarrow{\text{barrier method}} \quad \min \ f(x) - t\underbrace{\log x}_{\text{barrier function}} \triangleq g_t(x) \quad \text{s.t.} \ Ax=b$$

$t>0$

$\sum_i \log x_i$

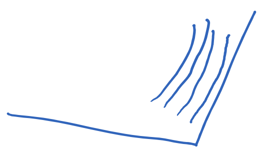
can transform $\tilde{A}z \leq \tilde{b}$ to $Ax=b$, $x\geq 0$

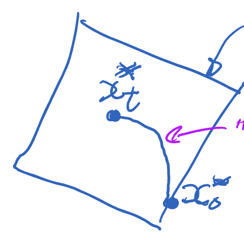
$\tilde{a}_i^T z \leq \tilde{b}_i$

transform

$\tilde{a}_i^T z + \xi_i = \tilde{b}_i \quad \xi_i \geq 0$

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





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

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

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

"self-concordant analysis"

IPM: # of Newton iterations is $\log(\frac{1}{\varepsilon})$ · some constant overall convergence

→ dimensionality

is appearing; but not condition # of f (vs. gradient method)

cost per iteration is $\approx O(d^3)$

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

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

Lagrangian dual fct.

Fenchel conjugate of f $\boxed{f^*(y) \triangleq \sup_x \langle y, x\rangle - f(x)}$

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

Fenchel duality: $A: \mathbb{R}^d \to \mathbb{R}^p$

$$p^* = \inf_x \; f(x) + g(Ax)$$

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

$\left(\frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_{i=1}^n H_i(w)\right)$

$\max_{\tilde{y}} -\ell_i(\tilde{y}) + \langle w, \psi_i(\tilde{y})\rangle$

note: not unique?

saddle pt. perspective for Lag. duality:

Lagrangian multiplier

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

"Lagrangian fct."

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

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

$$p^* = \inf_x \sup_{\alpha \geq 0} \mathcal{L}(x, \alpha)$$

"weak duality"

$$\geq \sup \;\; \inf \; \mathcal{L}(x, \alpha) \triangleq d^*$$

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

"weak duality"

Lagrangian dual fct.

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

note: $d(\alpha)$ always concave even if $f(x)$ is not concave

KKT conditions: → necessary conditions for x^* & α^* to be primal & dual optimums when f & h_i are differentiable fcts

KKT conditions

- x^* primal feasible (i.e. $h_i(x^*) \leq 0 \ \forall i \in I$)
- α^* dual feasible ($\alpha^* \geq 0$)
- $\nabla f(x^*) + \sum_i \alpha_i^* \nabla h_i(x^*) = 0$ [i.e. (x^*, α^*) station. point of $\mathcal{L}(x,\alpha)$]
- $\alpha_i^* h_i(x^*) = 0 \ \forall i$ complementary slackness

$h_i(x) \leq 0$

$\nabla h_i(x^*)$

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

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

(thus suppose get α^* which solves dual; then get x^* through KKT ⇒ x^* solves primal)

14h37

- $f(x) \geq d(\alpha)$ $\forall x$ & α feasible

  $\mathcal{L}(x,\alpha) \leq f(x)$ when α is feasible and x "

  $d(\alpha) \leq \mathcal{L}(x,\alpha) \leq f(x)$ $\forall x$

- $d(\alpha)$ is always concave in α
- $d(\alpha)$ is obtained through <u>unconstrained</u> opt.
- counter example for strong duality (see Boyd's book exercise 5.21)

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

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

- no unique "Lagrangian dual problem"

  $\rightarrow$ depends how you write the primal problem

SVM struct dual problem

cons. primal

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

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

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

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

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \Rightarrow \boxed{\sum_{\tilde{y} \in \mathcal{Y}_i} \alpha_i(\tilde{y}) = 1}$$

i.e. $\alpha_i \in \Delta_{|\mathcal{Y}_i|}$

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

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

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

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

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

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

$\dim(\alpha) = m$

$b \in \mathbb{R}^m$

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

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

dual problem: $$\boxed{\max_{\substack{\alpha_i\in\Delta_{|\mathcal{Y}_i|} \\ \text{simplex}}} -\frac{\lambda}{2}\|w(\alpha)\|^2 + \langle b,\alpha\rangle}$$

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

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

(exponentially many dual variables)

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

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

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

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

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

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

(*) <u>sparse</u> solution α^*

<u>primal-dual gap</u>:

$$\text{gap}(\alpha) = p(w(\alpha), \xi(\alpha)) - d(\alpha) \geq 0$$

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

certificate of suboptimality (*)

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

$$= \frac{1}{n}\sum_{i=1}^n \left[H_i(w) - \sum_{\tilde{y}} \alpha_i(\tilde{y})\, \ell_i(\tilde{y}) \right] + \lambda \langle w, \underbrace{w(\alpha)}_{\frac{1}{\lambda n}\sum_{i,\tilde{y}} \alpha_i(\tilde{y})\psi_i(\tilde{y})} \rangle$$

$$\boxed{\text{gap}(\alpha) = \frac{1}{n}\sum_{i=1}^n \Big[\underbrace{H_i(\overset{w(\alpha)}{w})}_{\max_{\tilde{y}} H_i(\tilde{y}; w)} - \sum_{\tilde{y}} \alpha_i(\tilde{y})\, \underbrace{H_i(\tilde{y}; \overset{w(\alpha)}{w})}_{\ell_i(\tilde{y}) - w^T\psi_i(\tilde{y})} \Big]}$$

$\leftarrow$ use to get a bound on dual suboptimality of α

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

a bound on suboptimality of α

and primal " of $w(\alpha)$