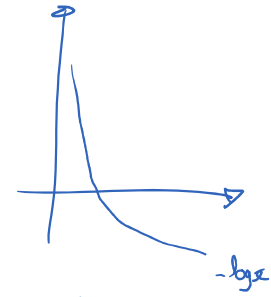


today: • I.P.M
• SVM dual



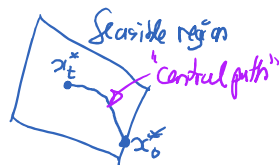
interior point method:

$$\begin{aligned} \min f(x) \\ \text{s.t. } Ax=b \\ x \succ 0 \end{aligned}$$

barrier method

$$\begin{aligned} \min f(x) + t \left(-\sum \log x_i \right) \triangleq g_t(x) \\ \text{s.t. } Ax=b \\ x_i > 0 \end{aligned}$$

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



$$x_\epsilon^* = \arg \min_{\text{s.t. } Ax=b} g_t(x)$$

interior pt. method: run Newton's method with warm start on $g_t(x)$ and decrease t with geometric schedule "self-concordant analysis"

convergence: Newton's method take $\log(\log(\frac{1}{\epsilon})) + \text{const.} \cdot (f(x_0) - f^*)$ iterations to reach ϵ -accuracy

IPM: ^{total} # of Newton iterations is $\log(\frac{1}{\epsilon})$. same constant overall convergence

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

Review of Lagrangian duality

primal problem

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

dual problem

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

↳ Lagrangian dual Sol.

Fenchel conjugate of f

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

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

Fenchel duality:

$$p^* = \inf_x f(x) + g(Ax) \quad A: \mathbb{R}^d \rightarrow \mathbb{R}^p$$

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

$\frac{\lambda \|\omega\|^2}{2} + \sum_{i=1}^n h_i(\omega)$
↓
max $\langle \lambda, y \rangle$
f $\langle \omega, \rho_i(y) \rangle$

note: not unique

saddle pt. necessary for dual

.saddle multipliers

1 | note: not unique

saddle pt. perspective for Lag. duality

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

Lagrange multiplier

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

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

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

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

"weak" duality

Lagrangian dual fct.

note: $d(x)$ always concave

even if $f(x)$ is not convex

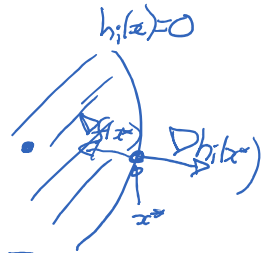
inf sup \geq sup inf?

$$\sup_{\alpha} L(x, \alpha) \geq \inf_{\alpha} L(x, \alpha)$$

Strong duality $p^* = d^*$

KKT conditions: \rightarrow necessary conditions for x^* & α^* to be primal & dual optimum resp. when f & h_i are diff. fcts.

- 14h44
- 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^*) stationary pt. of $L(x, \alpha)$ with respect to x]
 - $\alpha_i^* h_i(x^*) = 0 \forall i$ complementary slackness



* 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

$\exists x$ rel'int (dom f)
 s.t. $h_i(x) < 0$
 when h_i is not affine

(thus suppose get α^* which solves dual, then get x^* through KKT $\Rightarrow x^*$ sole primal)

\hookrightarrow see counter-example for strong duality (see Boyd's book ex. 5.21)

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

s.t. $x^2 \leq 0$
 y

(violates Slater's cond.)
and no strong duality

Other comments:

- $f(x) \geq d(\alpha)$ for α feasible
- $d(\alpha)$ is always concave in α
- $d(\alpha)$ is obtained through uncons = opt.
- no unique "separation dual problem" \rightarrow depends how you write primal problem

e.g. $h_1(x) \leq 0$ vs. $ah_1(x) \leq 0 \forall a > 0$

SVM strict dual problem

convex primal: $\min_{w, \xi} \frac{\lambda \|w\|^2}{2} + \frac{1}{n} \sum \xi_i$
 s.t. $\frac{1}{n} \xi_i \geq \frac{1}{n} H_i(\tilde{y}; w) \forall \tilde{y} \in \mathcal{Y}$
 $\hookrightarrow \ell_i(\tilde{y}) - w^T \varphi_i(\tilde{y})$

$$L(w, \xi; \alpha) = \frac{\lambda \|w\|^2}{2} + \frac{1}{n} \sum \xi_i + \frac{1}{n} \sum_{\tilde{y} \in \mathcal{Y}} \alpha_i(\tilde{y}) \left[\underbrace{\ell_i(\tilde{y}) - w^T \varphi_i(\tilde{y})}_{h_i(\tilde{y}; w, \xi)} - \xi_i \right]$$

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

$$\nabla_w L = 0 \Rightarrow \lambda w - \frac{1}{n} \sum_{\tilde{y} \in \mathcal{Y}} \alpha_i(\tilde{y}) \varphi_i(\tilde{y}) = 0$$

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

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

notation $\alpha \in \Delta_{|\mathcal{Y}_1|} \times \Delta_{|\mathcal{Y}_2|} \dots \times \Delta_{|\mathcal{Y}_n|}$
 $b = (b_i)_{i=1}^n \quad b_i(\tilde{y}) \triangleq \frac{1}{n} \ell_i(\tilde{y})$

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

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

$\dim(\alpha) = m$
 $b \in \mathbb{R}^m$
 $A \in \mathbb{R}^{d \times m}$

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

dual problem:

$$\max_{\alpha_i \in \mathcal{A}_i, 1 \leq i \leq n} \frac{-\Delta \|w(a)\|^2 + \langle b, a \rangle}{2}$$

(exponentially many dual variables)

$D \in \mathbb{R}^{m \times m}$

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

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

$$A \triangleq \left[\dots \frac{r_i(\tilde{y})}{\lambda_i} \dots \right]$$

\tilde{y} (column)

complementary slackness: $\alpha_i^*(\tilde{y}) > 0 \Rightarrow h_i(\tilde{y}; w^*) = 0$

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

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

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

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

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