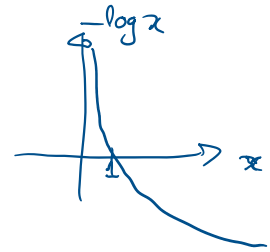
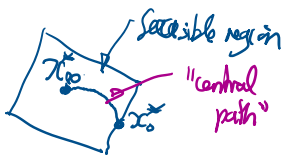


- today:
- IPM
 - duality
 - SVM dual



Interior point method:

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



barrier method $t \geq 0$

$$\min f(x) + t \underbrace{\left(-\sum_i \log x_i \right)}_{\text{barrier fct.}} \triangleq g_t(x) \\ \text{s.t. } Ax=b$$

$$x_t^* \triangleq \arg \min_x g_t(x) \text{ s.t. } Ax=b \\ \text{dom}(g_t(x)) = \{x : x > 0\}$$

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

convergence of Newton's: Newton's method take $\log(\log(\frac{1}{\epsilon})) + \text{cst.} \cdot (g_t(x_0) - g_t(x_t^*))$ iterations
 "self-concordant analysis"
 to reach ϵ -accuracy there is no condition # of g_t appearing

IPM: total # of Newton iterations is $\log(\frac{1}{\epsilon}) \cdot \text{some constants}$
 \hookrightarrow dimensionality is appearing linearly but not the condition # of f
 (vs. gradient methods)

Review of Lagrangian duality

primal problem

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

dual problem

$$d^* \triangleq \sup_{\alpha \geq 0} d(\alpha) \\ d(\alpha) \triangleq \inf_x f(x) + \sum_i \alpha_i h_i(x)$$

Lagrangian dual fct.

Fenchel duality

First: Fenchel conjugate of f

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

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

Fenchel duality:

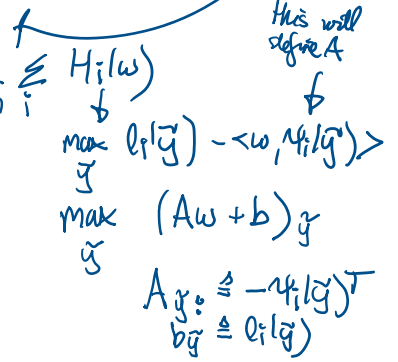
$$p^* = \inf_{x \in \mathbb{R}^d} f(x) + g(Ax) \quad A: \mathbb{R}^d \rightarrow \mathbb{R}^p$$

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

x

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

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



note: not unique!

saddle pt. perspective of Lag. duality:

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

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

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

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

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

"weak duality"

$\triangleq d(\alpha)$
Lagrangian dual fct.

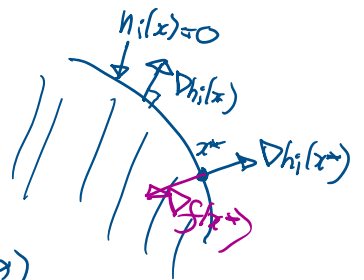
$$= \delta_C(x)$$

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

strong duality $p^* = d^*$ (have a "saddle pt.")

KKT conditions: \rightarrow necessary conditions for x^* & α^* to be primal & dual optimal (respectively) when f & h_i are differentiable fct.

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



if primal problem is convex and "Slater's condition" \Rightarrow strong duality 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

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

$\exists x$ in rel interior (dom f)
 s.t. $h_i(x) < 0 \forall i$
 when h_i is not affine

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

min $f(x, y)$
 s.t. $\frac{x^2}{y} \leq 0$

where $f(x, y) = \begin{cases} x^2/y & y > 0 \\ +\infty & \text{otherwise} \end{cases}$
 violates Slater's cond. and unbounded, no strong duality

other comments:

- $f(x) \geq d(\alpha) \forall x \text{ \& } \alpha$ 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 $\alpha > 0$

10h43

SVM struct dual problem

consider primal: min $\frac{\Delta \|w\|^2}{2} + \frac{1}{n} \sum_{i=1}^n \xi_i$

s.t. $\frac{1}{n} \xi_i \geq \frac{1}{n} H_i(\tilde{y}; w) \forall i \in \{1, \dots, n\}$
 $\hookrightarrow \xi_i(\tilde{y}) - w^T \psi_i(\tilde{y})$

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

$\frac{\partial L}{\partial \xi_i} = 0 \implies \sum_{\tilde{y}} \alpha_i(\tilde{y}) = 1$ (want $h_{i, \tilde{y}}(w, \xi) \leq 0$)
 i.e. $\alpha_i \in \Delta_{|\mathcal{Y}|}$

$\nabla_w L = 0 \implies \Delta w - \frac{1}{n} \sum_{i, \tilde{y}} \alpha_i(\tilde{y}) \psi_i(\tilde{y}) = 0$
 \rightarrow if $\alpha_i \in \Delta_{|\mathcal{Y}|}; d(\alpha) = +\infty$

$$\nabla_w L = 0 \Rightarrow Aw - \sum_{i, \tilde{y}} \alpha_i(\tilde{y}) \psi_i(\tilde{y}) = 0$$

$$w(\alpha) = A\alpha \Rightarrow \boxed{w(\alpha) \triangleq \sum_{i, \tilde{y}} \alpha_i(\tilde{y}) \psi_i(\tilde{y})} \quad \text{KKT conditions}$$

(here $w^* = w(\alpha^*)$)

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

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

$b \in \mathbb{R}^m$
 $\alpha \in \mathbb{R}^m$ exponentially big
 $m = \sum_i |\mathcal{Y}_i|$

$$b = (b_i)_{i=1}^m \quad b_i(\tilde{y}) \triangleq \sum_{\tilde{y}} l_i(\tilde{y})$$

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

$$A \triangleq \left[\dots \underbrace{\sum_{\tilde{y}} \psi_i(\tilde{y})}_{\substack{\text{column} \\ \tilde{y} \in \mathcal{Y}_i}} \dots \right]$$

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

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

$\lambda w(\alpha)$ \otimes

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

dual problem: $\max_{\substack{\alpha_i \in \Delta_{\mathcal{Y}_i} \\ 1 \leq i \leq n}} \frac{\lambda}{2} \|A\alpha\|^2 + \langle b, \alpha \rangle$

(exponentially many dual variables)

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

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

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

by primal feasibility

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

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

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

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

α^* is sparse

(aside: $y^{(i)}$ has $\psi_i(y^{(i)}) = 0$)

→ for input examples i s.t. $\arg\max H_i(\tilde{y}; w^*) = \{y^{(i)}\}$

then $\alpha_i(y^{(i)}) = 1 \Rightarrow$ this input example
does not influence w^*

i.e. not a "support vector"