LECTURE 12 CMPS 242 F17

OPTIMIZATION THEORY

LAGRANGIANS

DUALITY

HOW APPLIED TO SUPPORTVECTOR MACHINES

MORE ON KERNELS

NO CONSTRAINTS:

GIVEN A FUNCTION f (DIFFERENTIABLE) DEFINED ON A DOMAIN $\Omega \subseteq \mathbb{R}^n$.

MINIMIZE $f(w)$ $w \in \Omega$

NECESSARY CONDITION FOR MINIMUM

$$\underset{n \times 1}{\frac{\partial f(w)}{\partial w}} = 0$$

SUFFICIENT CONDITION FOR MINIMUM

$$\frac{\partial f(w)}{\partial w} = 0 \quad \text{AND} \quad \underset{n \times n}{\frac{\partial^2 f(w)}{(\partial w)^2}} \text{ POSITIVE SEMI-DEFINITE}$$

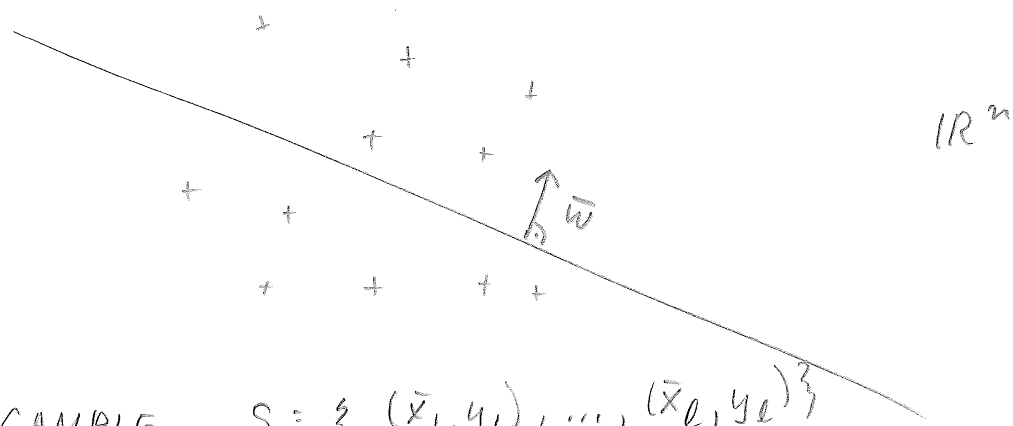
STRICT MIN. IF POSITIVE DEFINITE

$\underset{n \times n}{A}$ SYMMERIC POS. SEMI DEFINITE IF

1 - SYMMETRIC
2 - $\forall$ UNIT VECS $\bar{u}$ $\bar{u}^T A \bar{u} \geq 0$

or 2' - All EIGENVALS ≥ 0

EXAMPLE RIDGE REGRESSION
(LEAST SQUARES)



SAMPLE $S = \{(\bar{x}_1, y_1), \dots, (\bar{x}_\ell, y_\ell)\}$

$$f(w) = \frac{1}{\eta} \underbrace{\frac{1}{2}\|w - 0\|^2}_{\text{REGULARIZATION}} + \underbrace{\sum_{t=1}^{\ell} \frac{1}{2}(\bar{w}\cdot\bar{x}_t - y_t)^2}_{\text{LOSS}}$$

η IS TRADE-OFF PARAMETER

REMARK:

- $\eta \to \infty$ NO REGULARIZATION
  ONLY THE LOSS MINIMIZED
  - NOISE-FREE CASE: $\exists$ SOL. OF LOSS 0
    $w \cdot x_t = y_t$ WHICH SOLUTION TO CHOOSE?
  - WHAT IF NO SOLUTION OF LOSS 0?
- CLAIM: IF $\eta > 0$ THEN ALWAYS EXISTS UNIQUE SOLUTION

NOTE: NON-HOMOGENEOUS HP IS HOMOGENEOUS HP IN ONE DIM. LARGER

$w \cdot x + b = 0$ IFF $(w, b) \cdot (x, 1) = 0$

FIND MINIMUM BY SETTING DERIV'S TO ZERO:

$$\frac{\partial f(w)}{\partial w_i} = \frac{1}{\eta} w_i + \sum_{t=1}^{\ell} x_{t,i} \, (w \cdot x_t - y_t)$$

OR $$\frac{\partial f(w)}{\partial w} = \frac{1}{\eta} w + \sum_{t=1}^{\ell} x_t \, (x_t^T w - y_t)$$

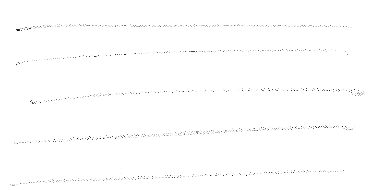
$$= \left(\frac{1}{\eta} I + \sum_{t=1}^{\ell} x_t x_t^T\right) w - \sum_{t=1}^{\ell} x_t y_t$$

$$\left.\frac{\partial f(w)}{\partial w}\right|_{w = w^*} = 0$$

$$w^* = \left(\frac{1}{\eta} I + \sum_{t=1}^{\ell} x_t x_t^T\right)^{-1} \sum_{t=1}^{\ell} x_t y_t$$

IN MATRIX NOTATION:

$$f(w) = \frac{1}{\eta} \frac{1}{2} \|w\|^2 + \frac{1}{2} \| \underbrace{X^T w - y}_{\text{RESIDUAL}} \|^2$$

$X =$  x_t

$\ell \times n$

$$\frac{\partial f(w)}{\partial w} = \frac{1}{\eta} w + X^T (Xw - y)$$

$$= \frac{1}{\eta} w + X^T X w - X^T y$$

$$= \left(\frac{1}{\eta} I + X^T X\right) w - X^T y$$

$$w^* = \left(\frac{1}{\eta} I + X^T X\right)^{-1} X^T y$$

$$\frac{\partial^2 f(w)}{(\partial w)^2} = \left(\frac{1}{\eta} I + X^T X\right)$$

STRICTLY POS. DEF. WHEN $\eta > 0$ & FINITE

SO STRICT MIN. AT W^*

$$\underbrace{\lim_{\eta \to \infty} \left(\frac{1}{\eta} I + X^T X\right)^{-1} X^T}_{\text{CONVERGES TO } X^+}$$

PSEUDO INVERSE

$w^* = X^+ y$ is SHORTEST SOLUTION

HESSIAN STRICTLY POS DEFINITE
NOT NECESSARY CONDITION FOR STRICT MINIMA

$f(x) = x^4$

$f'(x) = 4x^3$

$f''(x) = 12x^2$

MINIMIZE A FUNCTION f SUBJECT TO

EQUALITY CONSTRAINTS

GIVEN FUNCTIONS f, h_i $(1 \le i \le m)$
DEFINED ON A DOMAIN $\Omega \subseteq \mathbb{R}^n$

MINIMIZE $f(w)$, $\quad w \in \Omega$
SUBJECT TO: $h_i(w) = 0 \quad i = 1, \dots, m$

f OBJECTIVE FUNCTION
h_i CONSTRAINTS

LAGRANGIAN:

$$L(w, \beta) = f(w) + \sum_{i=1}^{m} \beta_i h_i(w)$$

$$= f(w) + \underbrace{\beta^T}_{1 \times m} \underbrace{h(w)}_{m \times 1}$$

W PRIMAL VARIABLES
β LAGRANGIAN OR DUAL VARS.
(ONE PER CONSTRIANT)

NECESSARY CONDITION

w^* MIN OF $f(w)$ SUBJECT TO = CONSTRAINTS

THEN $\frac{\partial L(w^*, \beta^*)}{\partial w} = 0$

$\frac{\partial L(w^*, \beta^*)}{\partial \beta} = 0$ "PRIMAL CONSTRAINTS"

FOR SOME VALUES β^* OF DUAL VARS

If $L(w, \beta^*)$ CONVEX IN w THEN
ABOVE CONDITION ALSO SUFFICENT

EXAMPLE: FIND LARGEST VOL. BOX WITH
SURFACE AREA c

MIN : $-wuv$
SUBJ. TO : $wu + uv + vw = c/2$

$$L(u, v, w, \beta) = -wuv + \beta(wu + uv + vw - \tfrac{c}{2})$$

$$\frac{\partial L}{\partial w} = -uv + \beta(u+v) = 0$$

$$\frac{\partial L}{\partial u} = -vw + \beta(v+w) = 0$$

$$\frac{\partial L}{\partial v} = -wu + \beta(w+u) = 0$$

$$\frac{\partial L}{\partial \beta} = \underbrace{wu + uv + vw - \frac{c}{2}}_{\text{PRIMAL CONSTRAINT}} = 0$$

ONLY ONE SOLUTION : $u^* = v^* = w^* = \sqrt{\frac{c}{6}}$, $\beta^* = \frac{1}{2}\sqrt{\frac{c}{6}}$

$\Rightarrow$ MAX VOL. IS CUBE

CHECK : SURFACE AREA $= 6 \cdot \sqrt{\frac{c}{6}}\sqrt{\frac{c}{6}} = c$

MAXIMUM ENTROPY DISTR.

H(p) :

-H(p) :

$$\text{MIN} \quad \overbrace{\sum p_i \ln p_i}^{-H(p)}$$

SUBJ. TO : $\sum_i p_i = 1$

$$L(\bar{p}, \beta) = \sum_i p_i \ln p_i + \beta\left(\sum_i p_i - 1\right)$$

$$\frac{\partial L}{\partial p_i} = \ln p_i + 1 + \beta = 0$$

$$p_i^* = e^{-1-\beta}$$

ALL p_i^* IDENTICAL

CONSTRAINT : $\sum p_i^* = 1$

IMPLIES : $p_i^* = \frac{1}{n}$, $\beta^* = -1 + \ln n$

PLUG IN

$$\sum_i \frac{1}{n} \ln \frac{1}{n} = -\sum_i \frac{1}{n} \ln n$$

$$= -\ln n$$

INEQUALITY CONSTRAINTS AS WELL

$$\left.\begin{array}{lll} \text{MINIMIZE} & f(w) & w \in \Omega \\ \text{SUBJECT TO} & g_i(w) \leq 0 & 1 \leq i \leq k \\ & h_i(w) = 0 & 1 \leq i \leq m \end{array}\right\} \text{PRIMAL}$$

LAGRANGIAN:

$$L(\underbrace{w}_{\text{PRIMAL VARS}}, \underbrace{\alpha, \beta}_{\text{LAGRANGIAN VARS OR DUAL VARS}}) = f(w) + \underset{k}{\alpha^T} g(w) + \underset{m}{\beta^T} h(w)$$

LAGRANGIAN DUAL PROBLEM

$$\left.\begin{array}{ll} \text{MAXIMIZE:} & \theta(\alpha, \beta) \\ \text{SUBJECT TO} & \alpha \geq 0 \\ \text{WHERE } \theta(\alpha, \beta) = \inf_{w \in \Omega} & L(\underbrace{w}_{\text{primal}}, \overbrace{\alpha, \beta}^{\text{dual}}) \end{array}\right\} \text{DUAL}$$

NOTE THAT β UNCONSTRAINT

PRIMAL:

MIN $\sum p_i \ln p_i$

$\sum p_i = 1$

WE SHOWED ALREADY THAT $p_i^* = \frac{1}{n}$

PLUGGING p^* INTO PRIMAL GIVES VALUE $-\ln n$

WHAT IS THE DUAL?

$$L(\bar{p}, \beta) = \sum p_i \ln p_i + \beta\left(\sum_i p_i - 1\right).$$

$$\frac{\partial L}{\partial p_i} = \ln p_i + 1 + \beta = 0$$

$$p_i^* = e^{-1-\beta} \quad \text{OPTIMAL PRIMAL}$$

PLUG INTO LAGR.

$$L(\bar{p}^*, \beta) = \sum_i e^{-1-\beta}(-1-\cancel{\beta}) + \cancel{\beta\left(\sum_i e^{-1-\beta}\right.} - 1\Big)$$

$$= -n e^{-1-\beta} - \beta$$

DUAL:

$$\max_{\beta} \underbrace{-n e^{-1-\beta} - \beta}_{\Theta(\beta)}$$

↑ UNCONSTR.

SOLVE DUAL FOR OPTIMUM β^*

$$\frac{\partial \Theta(\beta)}{\partial \beta} = -n e^{-1-\beta}(-1) - 1 = 0$$

$$e^{-1-\beta^*} = \frac{1}{n}$$

$$-1-\beta^* = -\ln n$$

$$\beta^* = (\ln n) - 1$$

$$\Theta(\beta^*) = -n e^{\cancel{-1} - ((\ln n) \cancel{-1})} - \ln n + 1$$

$$= -\cancel{n}\frac{1}{\cancel{n}} - \ln n \cancel{+1}$$

SAME VALUE AS PRIMAL

AN EXAMPLE WITH INEQUALITIES

$$\left.\begin{array}{ll} \text{MINIMIZE} & \frac{1}{2} w^T \underset{n \times n}{Q} w - \underset{n}{k}^T w \\ \text{SUBJECT TO} & \underset{m \times n}{X} w \leq \underset{m}{c} \end{array}\right\} \text{PRIMAL}$$

Q pos DEF.

$$L(w, \alpha) = \frac{1}{2} w^T Q w - k^T w + \alpha^T (Xw - c)$$

$$\frac{\partial L}{\partial w} = Qw - k + X^T \alpha = 0$$

$$w^* = Q^{-1}(k - X^T \alpha)$$

SUBSTITUTING INTO L GIVES

$$\frac{1}{2}\left(Q^{-1}(k - X^T\alpha)\right)^T \cancel{Q Q^{-1}}(k - X^T \alpha) - k^T Q^{-1}(k - X^T \alpha)$$

$$- \alpha^T (X Q^{-1}(k - X^T \alpha) - c)$$

$$\vdots$$

$$\left.\begin{array}{ll} \text{MAX} & -\frac{1}{2}\alpha^T P \alpha - \alpha^T d - \frac{1}{2} k^T Q k \\ \text{SUBJECT TO} & \alpha \geq 0 \end{array}\right\} \text{DUAL}$$

SIMPLER CONSTRAINTS

WHERE $P = X Q^{-1} X$
$d = c - X Q^{-1} k$

HOW ARE PRIMAL AND DUAL RELATED?

WEAK DUALITY TH:

$w \in \Omega$ FEASIBLE SOL. TO PRIMAL

(α, β) " " DUAL

THEN $\underset{\text{MINIMIZED}}{f(w)} \geq \underset{\text{MAXIMIZED}}{\theta(\alpha, \beta)}$

PROOF:

$$\theta(\alpha,\beta) = \inf_{\tilde{w} \in \Omega} L(\tilde{w}, \alpha, \beta)$$

$$\leq L(w, \alpha, \beta)$$

$$= f(w) + \underbrace{\alpha^T}_{\geq 0} \underbrace{g(w)}_{\leq 0} + \beta' \underbrace{h(w)}_{= 0}$$

FEASIBILITY OF α

FEASIBILITY OF w

$$\overset{(*)}{\leq} f(w)$$

$\square$

CONCLUSION:

VALUE OF DUAL UPPER BOUNDED BY VALUE OF PRIMAL

$\downarrow$ PUSH DOWN

$$\inf\{f(w) : g(w) \le 0,\ h(w) = 0\}$$

$$\vee\!\mid \qquad \text{DUALITY GAP}$$

$$\sup\{\theta(\alpha, \beta) : \alpha \geq 0\}$$

$\uparrow$ PUSH UP

WHEN THE GAP IS 0
THEN THIS GIVES FEASIBLE SOLUTIONS

If $f(w^*) = \theta(\alpha^*, \beta^*)$, WHERE
$\alpha^* \geq 0$ $g(w^*) \le 0$ $h(w^*) = 0$

THEN w^* & (α^*, β^*) SOLVE THE PRIMAL AND DUAL PROBLEM, RESP.

ALSO:
$\alpha_i^* g_i(w^*) = 0$ for $1 \le i \le k$

WHY? $(*)$ IS TIGHT IFF THE ABOVE HOLDS

FOR US THE CONSTRAINTS ARE ALWAYS AFFINE:

$$h_i(w) = (\bar{a}_i^T \bar{w} - b_i = 0)$$

$$g_i(w) = (\bar{c}_i^T w - d_i \leq 0)$$

ALSO f ALWAYS CONVEX

STRONG DUALITY TH:

ASSUME f CONVEX

FOR A CONVEX DOMAIN Ω

g_i, h_i AFFINE

THEN THE FOLLOWING OPTIM. PROBLEM

HAS DUALITY GAP 0

$$\begin{array}{lll} \text{MIN } f(w) & w \in \Omega & \\ \text{SUBJ. TO:} & g_i(w) \leq 0 & 1 \leq i \leq k \\ & h_i(w) = 0 & 1 \leq i \leq m \end{array}$$

A TH. YOU SHOULD REMEMBER

MIN. $f(w)$ FOR $w \in \Omega$

SUBJ. TO: $g_i(w) \leq 0 \quad 1 \leq i \leq k$

$h_i(w) = 0 \quad 1 \leq i \leq m$

Ω CONVEX, g_i, f_i AFFINE

NEC. & SUFF. CONDITIONS FOR w^* TO BE AN OPTIMUM ARE EXISTANCE OF α^*, β^* S.T.

$$\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial w} = 0$$

$$h_i(w^*) = 0 \quad 1 \leq i \leq m$$

$$g_i(w^*) \leq 0 \quad 1 \leq i \leq k$$

$$\alpha_i^* g_i(w^*) = 0 \quad 1 \leq i \leq k$$

$$\alpha_i^* \geq 0$$

KARUSH-KUHN-TUCKER CONDITIONS

CONSTR. $g_i(w^*) \leq 0$ ACTIVE IF $g_i(w^*) = 0$

INACTIVE < 0

ACTIVE $\Rightarrow \alpha_i^* \geq 0$

INACTIVE $\Rightarrow \alpha_i^* = 0$