Lecture 14 **EE 660** Oct 4, 2018

Announcements

- Midterm ground rules
  - Closed book with 1 formula sheet allowed (2-sided, 8.5" x 11")
  - A simple or scientific calculator (single-purpose device, non-programmable) is allowed. Note that software-based calculators (e.g., on a smartphone) are not allowed.
- Project Assignment has been posted
  - See its timeline for due dates
- HW7 (Project Proposal) has been posted
  - Due Friday, 10/12
  - If you choose to post your choice of Kaggle competition topic on piazza, your post is due Monday, 10/8, 12:00 noon.

Today's Lecture

- Three types of projects - choose one
- Regularization (AML view)

CLASS PROJECT

3 TYPES:

(1) PROJECT OF YOUR OWN DESIGN, BASED ON REAL-WORLD DATA.

> FIND / CHOOSE A DATASET.

> SET GOALS.

> DESIGN PROJECT, AND DEVELOP APPROACH.

(2) COLLABORATIVE PROJECT BASED ON A CURRENT KAGGLE COMPETITION.

> FIND / CHOOSE A KAGGLE COMPETITION.

> SEE IF THERE ARE OTHER TEAMS / INDIVIDUALS IN CLASS TO WORK ON THE SAME ~~COM~~ COMPETITION.

→ (OPTIONAL) POST YOUR CHOICE ON PIAZZA – BY MONDAY 10/8.

(3) PROJECT BASED ON EXPERIMENTAL OR THEORETICAL INVESTIGATION.

> DEFINE SOME QUESTION(S).

> DESIGN NUMERICAL EXPERIMENTS OR THEORETICAL APPROACH TO ANSWER QUESTION(S).

REGULARIZATION AND COMPLEXITY (AML VIEW)

VC BOUND VIEW (FROM BEFORE):

$$E_{out}(h) \leq E_{in}(h) + \Omega(\mathcal{H}, N, \delta) \quad \forall h \in \mathcal{H}$$

LEARNING ALG. FINDS:

$$h_g = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \, E_{in}(h)$$

$\Omega(\mathcal{H})$ DEPENDS ON $\mathcal{H}$ BUT NOT ON h_g.

WITH REGULARIZATION ($\underline{w} = \underline{w}^{(0)}$ NON-AUGM.)

LET $f_{obj}^{(r)}(h_{\underline{w}}) = E_{in}(h_{\underline{w}})$ SUBJECT TO

$$\underline{w}^T \underline{w} \leq C$$

(C IS A PARAMETER).

(i) $h_g = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \, E_{in}(h_{\underline{w}})$ SUBJECT TO $\underline{w}^T \underline{w} \leq C$.

EQUIVALENT TO:

(ii) $h_g = \underset{h \in \mathcal{H}'}{\operatorname{argmin}} \; E_{in}(h_{\underline{w}})$

in which $\mathcal{H}' = \{ h \mid h \in \mathcal{H} \text{ and } \underline{w}^T \underline{w} \le C \}$

→ DIFFERENT HYP. SET.

$\Rightarrow d_{vc}(\mathcal{H}') \le d_{vc}(\mathcal{H})$.

GRAPHICAL VIEW

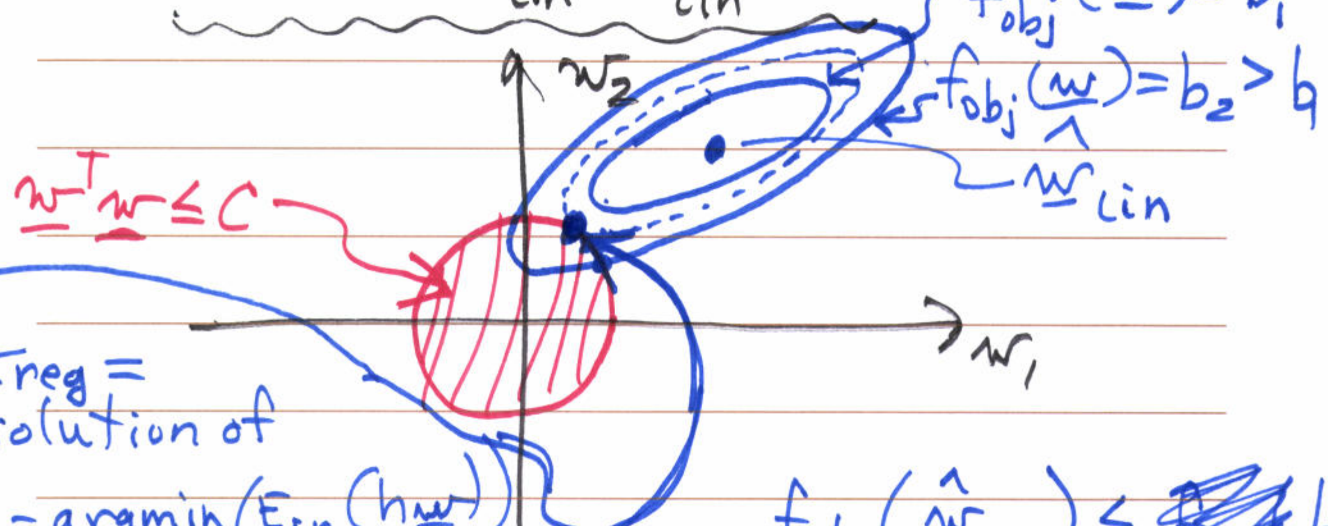
$f_{obj}(\underline{w}) \triangleq E_{in}(h_{\underline{w}}) = \frac{1}{N}\sum_{n=1}^{N} (\underline{w}^T \underline{x}_n + w_0 - y_n)^2$

example

$= \text{min. at } \hat{\underline{w}}_{lin}$

USING (i) ABOVE:

CASE I: $\hat{\underline{w}}_{lin}^T \hat{\underline{w}}_{lin} > C$



$\hat{\underline{w}}_{reg}$ = solution of

$h_g = \operatorname{argmin}(E_{in}(h_{\underline{w}}))$

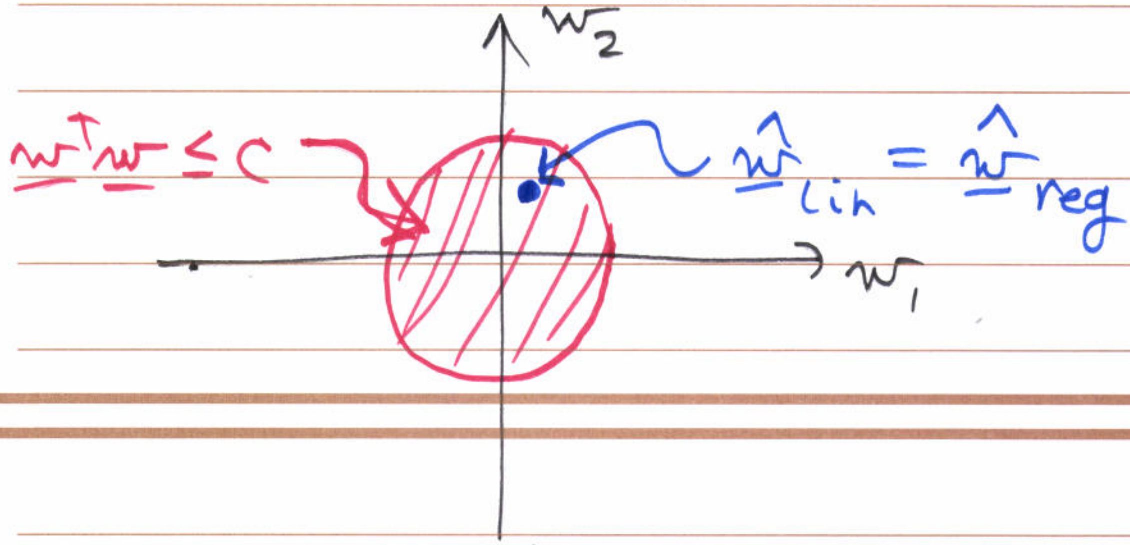
s.t. $\underline{w}^T \underline{w} \le C$

$f_{obj}(\hat{\underline{w}}_{lin}) < b_1$

→ IN THIS CASE, $\underline{\hat{w}}_{reg}^T \underline{\hat{w}}_{reg} = C$,

AND $\underline{\hat{w}}_{reg} \neq \underline{\hat{w}}_{lin}$.

CASE 2: ~~$\underline{\hat{w}}_{reg}$~~ $\underline{\hat{w}}_{lin}^T \underline{\hat{w}}_{lin} \leq C$:



→ IN THIS CASE, $\underline{\hat{w}}_{reg}^T \underline{\hat{w}}_{reg} \leq C$, AND

$\underline{\hat{w}}_{reg} = \underline{\hat{w}}_{lin}$.

~~USE~~

USE LAGRANGIAN OPTIMIZATION

(WITH INEQUALITY CONSTRAINT):

$$L(\underline{w}, \lambda_c) = E_{in}(h_{\underline{w}}) - \lambda_c \left[C - \underline{w}^T \underline{w} \right],$$

$$\lambda_c \geq 0.$$

$$\nabla_{\underline{w}} L = \nabla_{\underline{w}} \left\{ E_{in}(h_{\underline{w}}) - \lambda_c \left[\cancel{C} - \underline{w}^T \underline{w} \right] \right\} \Big|_{\underline{w} = \underline{w}_{reg}} = 0$$

$$\nabla_{\underline{w}} \left\{ E_{in}(h_{\underline{w}}) + \lambda_c \underline{w}^T \underline{w} \right\} \Big|_{\underline{w} = \underline{w}_{reg}} = 0$$

$\Rightarrow$ $\underline{w}_{reg}$ (LOCALLY) MINIMIZES:

$$E_{aug}(h_{\underline{w}}) \triangleq E_{in}(h_{\underline{w}}) + \lambda_c \underline{w}^T \underline{w}, \quad \lambda_c \geq 0.$$

TREAT λ_c AS A PARAMETER.

THUS,

$$\text{(iii)} \quad h_g = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \left\{ E_{in}(h_g) + \lambda_c \underbrace{\underline{w}^T \underline{w}}_{} \right\}, \quad \lambda_c \geq 0.$$

ℓ_2 REGULARIZER

(ALSO CALLED "WEIGHT DECAY").

(i), (ii) HAD C AS PARAMETER.
(iii) HAS λ_C AS PARAMETER.

~~SUMMARY:~~ THUS, WE CAN WRITE:

$$E_{aug}(h_{\underline{w}}) = E_{in}(h_{\underline{w}}) + \lambda \underline{w}^T \underline{w}, \quad \lambda \geq 0$$

(UNCONSTRAINED OPTIMIZATION PROBLEM)

AND MORE GENERALLY:

$$E_{aug}(h_{\underline{w}}) = E_{in}(h_{\underline{w}}) + \frac{\lambda'}{N} \Omega(h_{\underline{w}}),$$

$$\lambda' \triangleq N\lambda \geq 0$$

[AML FIG. 4.5]

[AML FIG. 4.7]