Lecture 17 **EE 660** Oct 16, 2018

Announcements

- HW 8 is due Friday
- Sample midterm exam will be posted
- Discussion 9 will work through example problems, in review for midterm
- Lecture ~~17~~ 18 will include review of lecture material for midterm

---

Today's topics

- Validation and test

TEST AND VALIDATION

KEY RELATIONS AND BOUNDS

$$(i)\ \left\{\begin{array}{l} \left.\begin{array}{l} E_{out}(h_g) \le E_{\mathcal{D}_0}(h_g)+\varepsilon_{eff} \\ E_{out}(h_g) \le E_{\mathcal{D}_0}(h_g)+\varepsilon_{vc} \end{array}\right\} \text{with probability } \le 1-\delta \\ \varepsilon_{eff} = \sqrt{\frac{8}{N}\ln\frac{4\, m_{\mathcal{H}}(2N)}{\delta}} \\ \varepsilon_{vc} = \sqrt{\frac{8}{N}\ln\frac{4\left[(2N)^{d_{vc}}+1\right]}{\delta}} \\ \varepsilon_{eff} \le \varepsilon_{vc}, \quad N = |\mathcal{D}_0| \end{array}\right.$$

e.g.: $\mathcal{D}_0 = \mathcal{D}_{Tr}$, $\mathcal{H} = \mathcal{H}_{Tr}$, $N = |\mathcal{D}_{Tr}|$,

$d_{vc} = d_{vc}(\mathcal{H}_{Tr})$, $m_{\mathcal{H}}(2N) = m_{\mathcal{H}_{Tr}}(2N)$.

$$(ii)\ \begin{cases} E_{out}(h_g) \le E_{\mathcal{D}_a}(h_g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \\ \quad \text{with pr.} \ge 1-\delta. \\ M = |\mathcal{H}|,\ N = |\mathcal{D}_a| \end{cases}$$

e.g.: if $\mathcal{D}_a = \mathcal{D}_{Test}$, then usu. $M = 1$.

BOTH (i) & (ii):

GIVE BOUNDS ON $E_{out}(h_g)$;

OR CAN BE USED TO COMPARE $E_{out}(h_i) \leftrightarrow E_{out}(h_j)$

BY COMPARING $E_{\mathcal{D}}(h_i) \leftrightarrow E_{\mathcal{D}}(h_j)$ WITH SIMILAR $\mathcal{E}_{eff}$, $\mathcal{E}_{VC}$, OR $\mathcal{E}_M$ (e.g., IF h_i AND h_j ARE FROM SAME HYPOTHESIS SET $\mathcal{H}$).

CONTEXT:

IN COMPARING MODELS, WE WANT TO COMPARE $E_{out}(h_i)$ WITH $E_{out}(h_j)$, OR EST.'S ON BOUNDS FOR THESE.

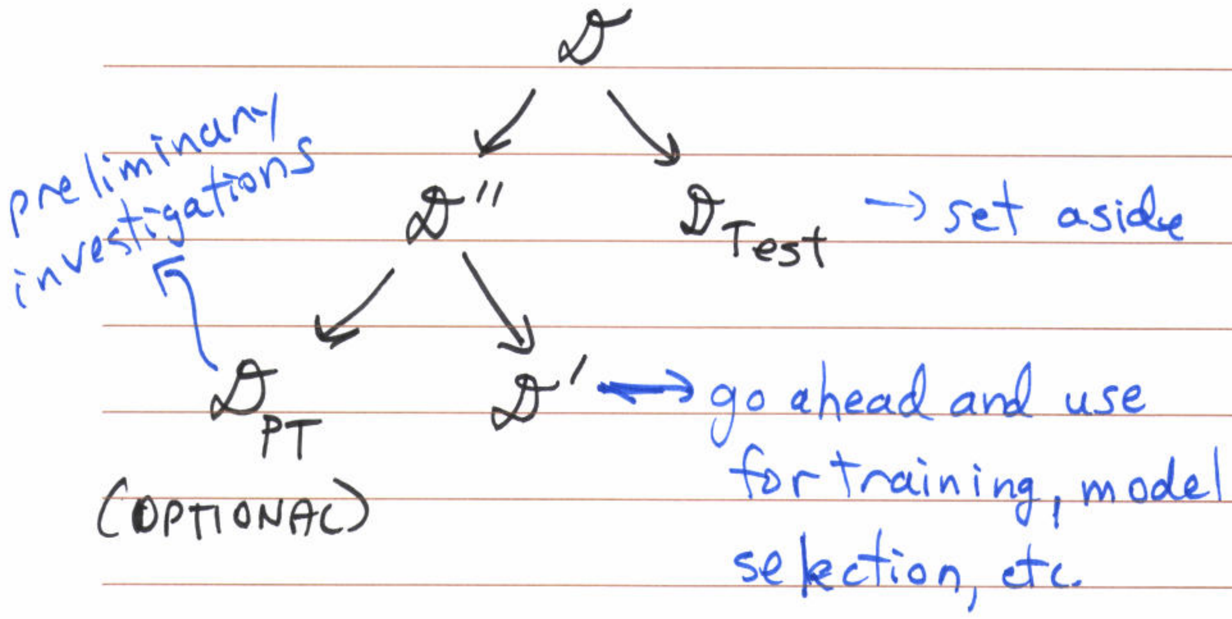
EX! FIND BEST VALUE FOR λ FOR A REGULARIZER.

$$f(\underline{w}, \lambda) = E_{in}(\underline{w}) + \lambda \|\underline{w}\|_b$$

$\min_{\underline{w}, \lambda} f$ TO PREDICT APARTMENT RENT.

TYPICAL TRAINING ALGORITHM (RIDGE REGRESSION) FINDS $\hat{\underline{w}}$ FROM $\mathcal{D}_{Tr}$, FOR A GIVEN λ. MODEL SELECTION FINDS λ.

WE HAVE DONE:



EX: IRIS CLASSIFICATION

SUPPOSE: $\hat{f}(\underline{x}) = \text{sign}\{\underline{w}^T\underline{x} + w_0\}$

$$\rightarrow \mathcal{H}' = \{ h_{\underline{w}, w_0}(\underline{x}) = \hat{f}(\underline{x}) \mid \underline{w} \in \mathbb{R}^D, w_0 \in \mathbb{R} \}$$

$$f_{obj}(\underline{w}, w_0, \lambda) = E_{in}(\underline{w}, w_0) + \lambda \|\underline{w}\|_1$$

MODELS $(\mathcal{H}', \lambda_1), (\mathcal{H}', \lambda_2), \cdots, (\mathcal{H}', \lambda_M)$
WITH $\lambda_m = m\,\delta_\lambda$, $m = 1, \cdots, M$.

FOR MODEL SELECTION, CAN USE A VALIDATION SET $\mathcal{D}_{val}$, SO:

$$\begin{tikzcd} \mathcal{D}' \arrow[r] \arrow[dr] & \mathcal{D}_{Tr} \\ & \mathcal{D}_{val} \end{tikzcd}, \quad \mathcal{D}_{Tr} \cap \mathcal{D}_{val} = \emptyset$$

PROCEDURE

1. $$\begin{tikzcd} & \mathcal{D}_{Test} \\ \mathcal{D} \arrow[ur] \arrow[r] \arrow[dr] & \mathcal{D}_{PT} \text{ (OPTIONAL)} \\ & \mathcal{D}' \end{tikzcd}$$

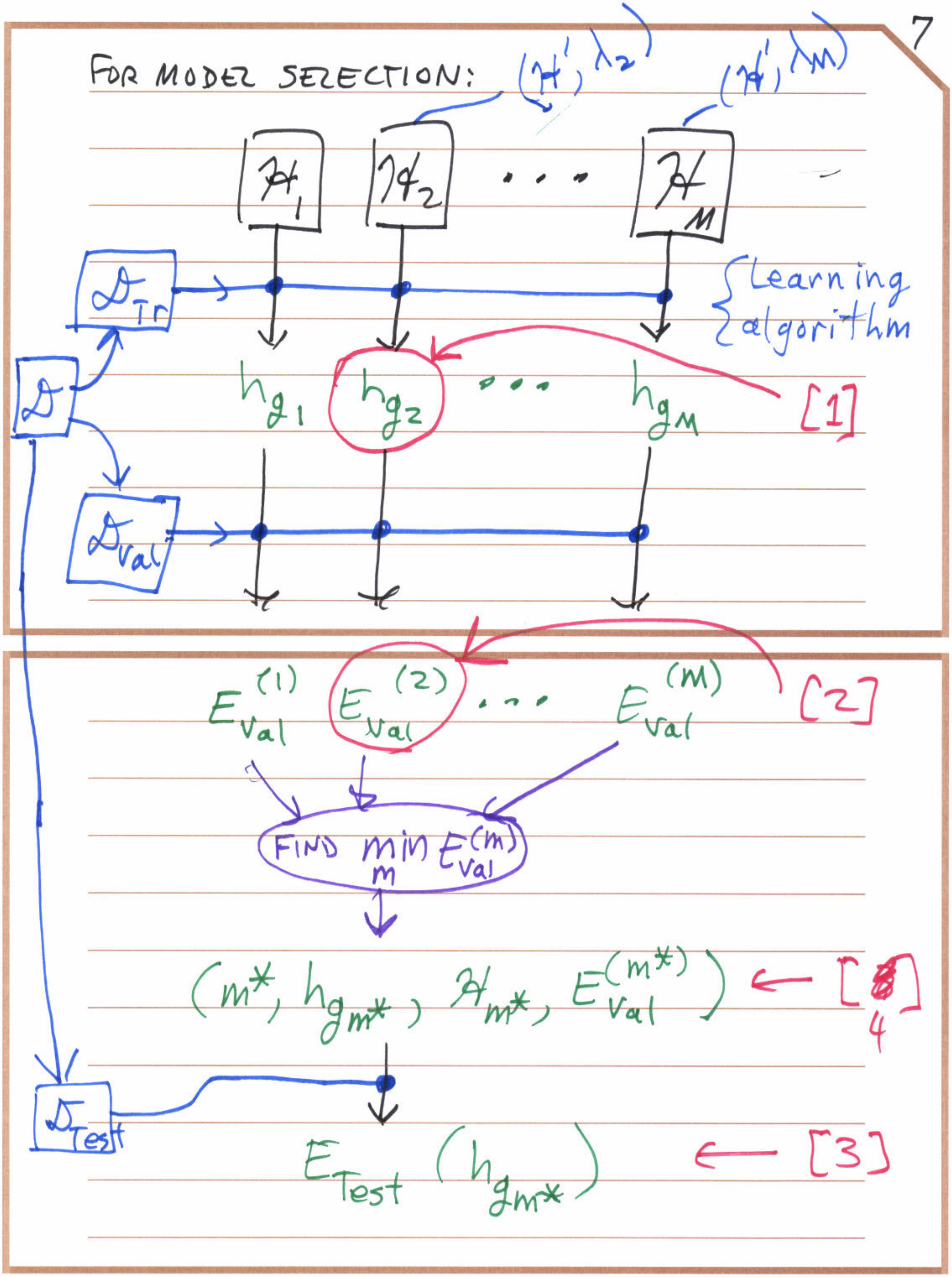
2. OPTIONALLY USE $\mathcal{D}_{PT}$.

3. SET UP $\mathcal{H}'$ AND MODELS:

$$\mathcal{H}_m \triangleq (\mathcal{H}', \lambda_m), \quad m=1, \cdots, M.$$

4. SEPARATE $$\begin{tikzcd} \mathcal{D}' \arrow[r] \arrow[dr] & \mathcal{D}_{Tr} \\ & \mathcal{D}_{val}. \end{tikzcd}$$

FOR MODEL SELECTION:



[1] WHAT HYPOTHESIS SET DID h_{g_2} COME FROM?
A: $\mathcal{H}_2$.
THEORETICAL BOUNDS $E_{out}(h_{g_2})$ USING $\mathcal{D}_{Tr}$?

→ USE (i) WITH $\mathcal{D}_0 = \mathcal{D}_{Tr}$,
$\mathcal{H} = \mathcal{H}_2$, $N = |\mathcal{D}_{Tr}|$
$d_{vc} = d_{vc}(\mathcal{H}_2)$, $m_{\mathcal{H}}(2N) = m_{\mathcal{H}_2}(2N)$

[2] WHAT HYP. SET DO WE USE TO GET $E_{out}(h_{g_2})$ BOUND, BASED ON $\mathcal{D}_{Val}$?

$\mathcal{H} = \{h_{g_2}\}$
→ USE (ii) WITH $\mathcal{D}_a = \mathcal{D}_{val}$, $M=1$.

[3] HOW EST. $E_{out}(h_{g_{m^*}})$ FROM $\mathcal{D}_{Test}$?
$\mathcal{H} = \{h_{g_{m^*}}\}$
→ USE (ii), WITH $\mathcal{D}_a = \mathcal{D}_{Test}$, $M=1$.

[4] How EST. $E_{out}(h_{g_{m^*}})$ FROM $\mathcal{D}_{val}$?

$\mathcal{H}'' = \{h_{g_1}, h_{g_2}, \dots, h_{g_M}\}$.

→ USE (ii), WITH $M = |\mathcal{H}''| = M$,

$\mathcal{D}_a = \mathcal{D}_{val}$.

OR → USE (i), WITH $m_{\mathcal{H}}(2N) = m_{\mathcal{H}''}(2N)$,

or $d_{vc}(\mathcal{H}) = d_{vc}(\mathcal{H}'')$, $\mathcal{D}_0 = \mathcal{D}_{val}$: