Thursday, 10/1/2020

# EE 660

# MACHINE LEARNING FROM SIGNALS: FOUNDATIONS AND METHODS

Prof. B. Keith Jenkins

## Lecture 12

## Announcements

- Homework 4 (Week 5) is due tomorrow.
- Homework 5 (Week 6) will be posted.

## Today's Lecture

- Overfitting (part 2)
- Regularization (AML view)

# Continue examples / experiments of overfitting (from last lecture)

Now consider:

Target fcn. $f(x)$ is $d=50$ order polynomial, no noise.

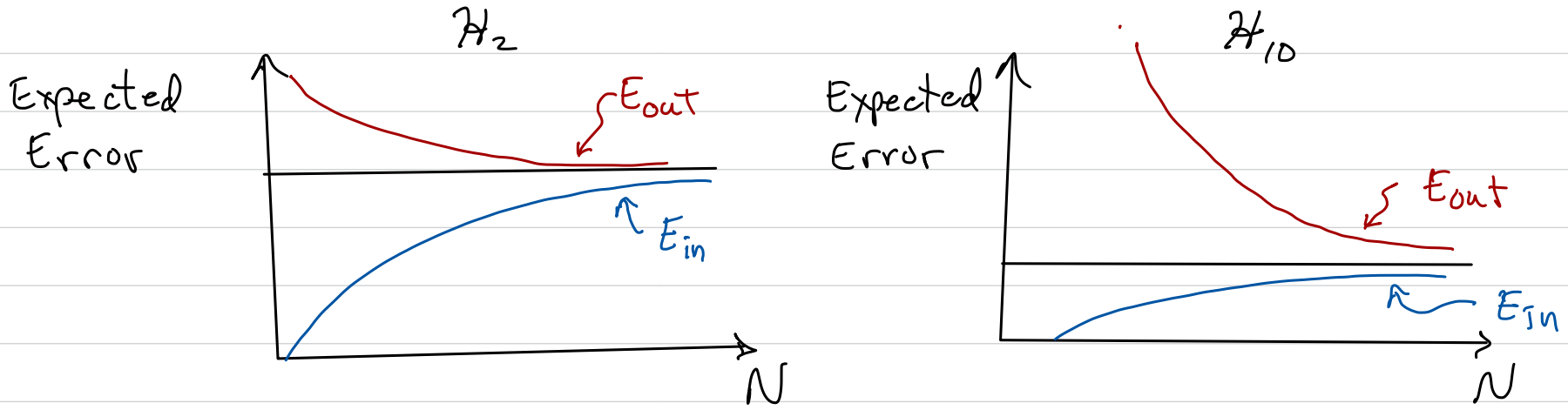
$\mathcal{H}_2$ vs. $\mathcal{H}_{10}$: $\mathcal{H}_{10}$ overfits! Even though $\mathcal{H}_{10}$ is much simpler than $f(x)$.

→ More data would help $\mathcal{H}_{10}$.

[AML fig. and table pp. 120-121]

⇒ Best hypothesis set complexity depends on quantity of data.

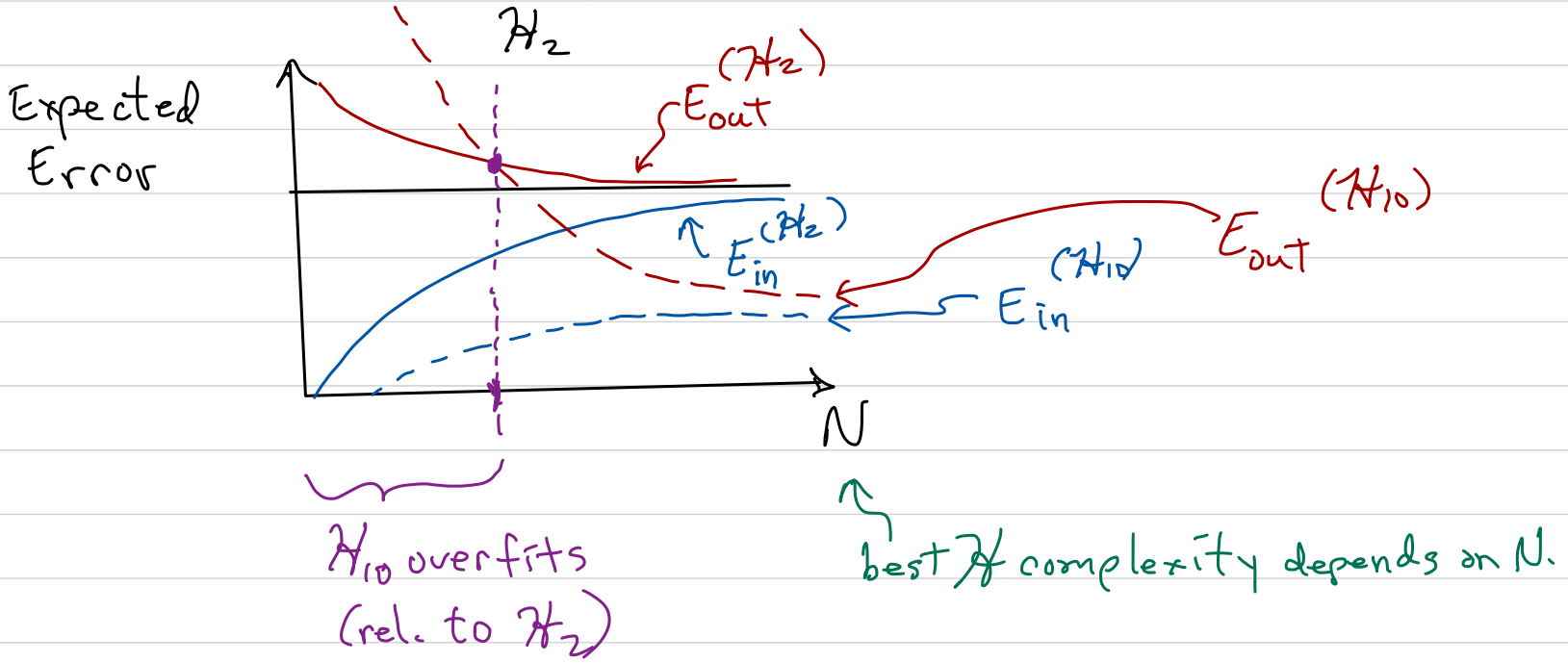
## Learning Curves




Expected Error: $\mathbb{E}_{\mathcal{D}}\{E(h_g^{(\mathcal{D})})\}$

If $E_{in}(h_k) \leq E_{in}(h_j)$, and $E_{out}(h_k) > E_{out}(h_j)$,

then $h_k$ overfits the data (relative to $h_j$).




What parameters affect amount of overfitting?

N, $\sigma$, complexity of $\mathcal{H}$, complexity of $f(\underline{x})$.

↑ noise on data

[A good example / experiment will be covered in Discussion 6.]

# Regularization and Complexity [AML 4.2]

VC bound view:

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

Learning alg. finds $h_g = \underset{h \in \mathcal{H}}{\operatorname{argmin}}\ E_{\mathcal{D}_{Tr}}(h)$

Note: $\Omega$ depends on $\mathcal{H}$ but not on $h_g$.

With regularization: $(\underline{w} = \underline{w}^{(0)})$

Let $f_{obj}^{(r)}(h_{\underline{w}}) = E_{\mathcal{D}}(h_{\underline{w}})$ subject to $\underline{w}^T\underline{w} \leq C$

"soft order constraint"

(i) $h_g = \underset{h \in \mathcal{H}}{\operatorname{argmin}}\ E_{\mathcal{D}}(h)$ subject to $\underline{w}^T\underline{w} \leq C$ ($C$ is a parameter)

equivalent to:

(ii) $h_g = \underset{h \in \mathcal{H}'}{\operatorname{argmin}}\ E_{\mathcal{D}}(h)$ in which $\mathcal{H}' = \{h \mid h \in \mathcal{H} \text{ and } \underline{w}^T\underline{w} \leq C\}$

$\mathcal{H}'$ is different than $\mathcal{H}$ (in general)

Note: $d_{vc}(\mathcal{H}') \leq d_{vc}(\mathcal{H})$.

Graphical view:

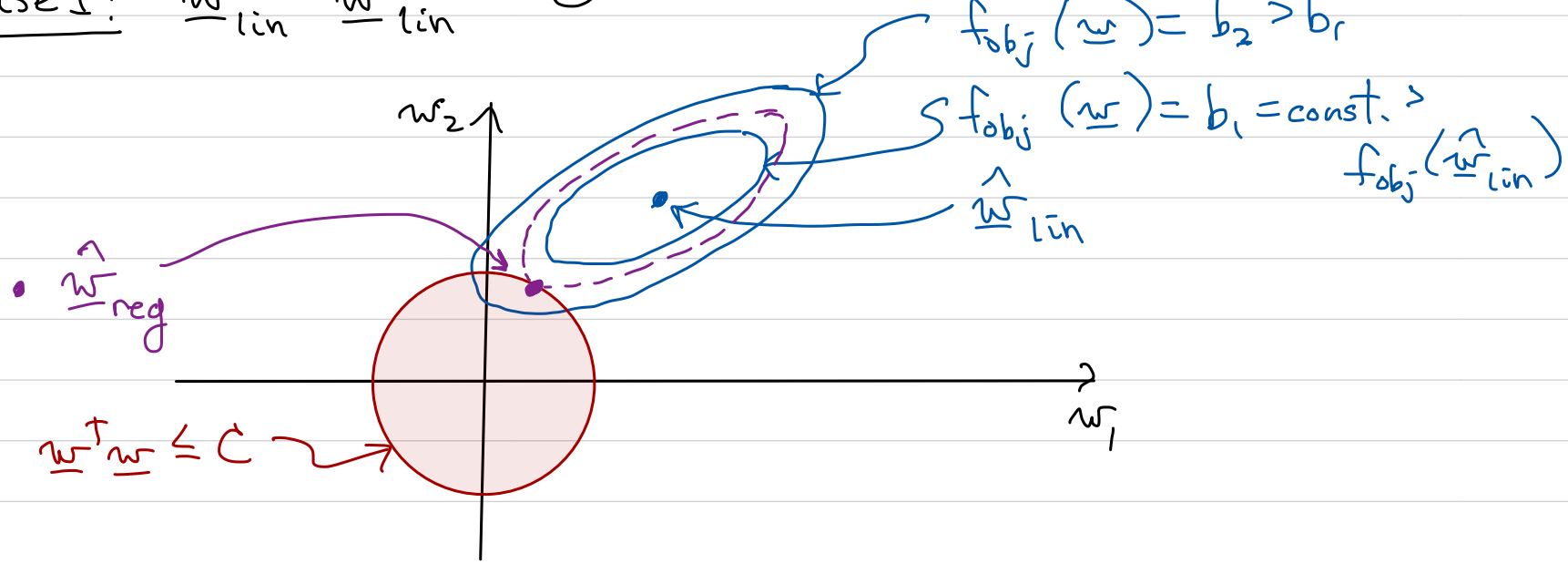
$$f_{obj}(\underline{w}) \triangleq E_{\mathcal{D}}(h_{\underline{w}}) \overset{\text{for example}}{=} \frac{1}{N}\sum_{n=1}^{N}\left(\underline{w}^T\underline{x}_n + w_0 - y_n\right)^2$$

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

(lin for linear)
(unconstrained optimum)

Using (i) above:

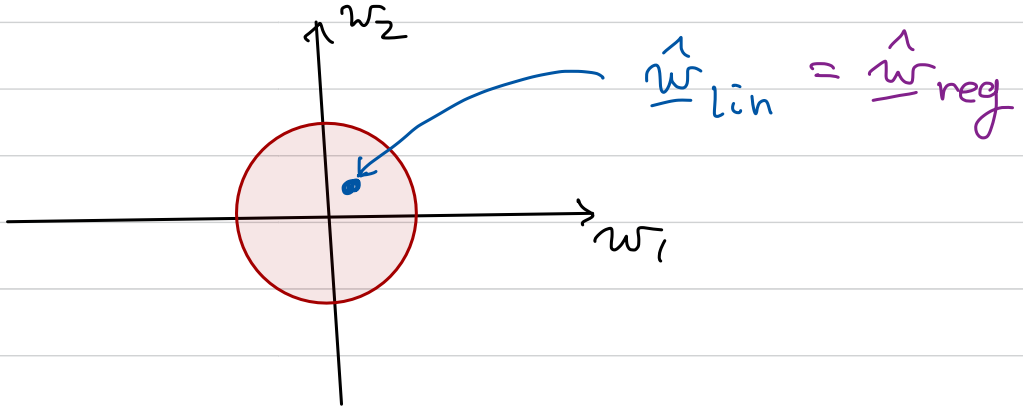
Case 1: $\hat{\underline{w}}_{lin}^T \hat{\underline{w}}_{lin} > C$



$\hat{w}_{reg}$ is the min. of $f_{obj}$ subject to $w^T w \leq C$.

→ In Case 1, $\hat{w}_{reg}^T \hat{w}_{reg} = C$, and $\hat{w}_{reg} \neq \hat{w}_{Lin}$.

Case 2 : $\hat{w}_{Lin}^T \hat{w}_{Lin} \leq C$




In Case 2, $\hat{w}_{reg}^T \hat{w}_{reg} \leq C$, and $\hat{w}_{Lin} = \hat{w}_{reg}$.