Announcements

- HW6 was due today
- Midterm ground rules vote to come
- Project Assignment is coming…
  - Dataset tips (v0.6) was posted

Today's Lecture

- Overfitting

OVERFITTING [AML 4.1]

DEF: OVERFIT IS "AN ANALYSIS WHICH CORRESPONDS TOO CLOSELY OR EXACTLY TO A PARTICULAR SET OF DATA".

[OXFORD DICTIONARY]

COMMON SYMPTOM OF OVERFITTING: PICKING A HYPOTHESIS WITH LOWER E_{in} RESULTS IN A HIGHER E_{out}.

EX: (EXPERIMENTS)

AML. FIGURES [4.1 (a) & PREVIOUS]

$f(x)$ = 10^{th} ORDER POLYN.

DATA = $f(x)$ + NOISE.

$N = 15$

2 HYPOTH. SETS: $\mathcal{H}_2$, $\mathcal{H}_{10}$.

$\downarrow$ $\downarrow$

h_{g_2} $h_{g_{10}}$

$\Rightarrow$ h_{g_2} FIT IS GOOD; $h_{g_{10}}$ OVERFITS!

$\Rightarrow$ NOISE DISTRACTED THE FIT.

$\therefore$ NOISE (QUALITY OF DATA) MATTERS, EVEN IF MODEL MATCHES THE TARGET FCN.

Now $f(\underline{x}) = 50^{\text{th}}$ ORDER POLYN.

DATA = $f(x)$ (NO NOISE)

$N = 15$

$\Rightarrow$ h_{g_2} FIT IS GOOD; $h_{g_{10}}$ OVERFITS.

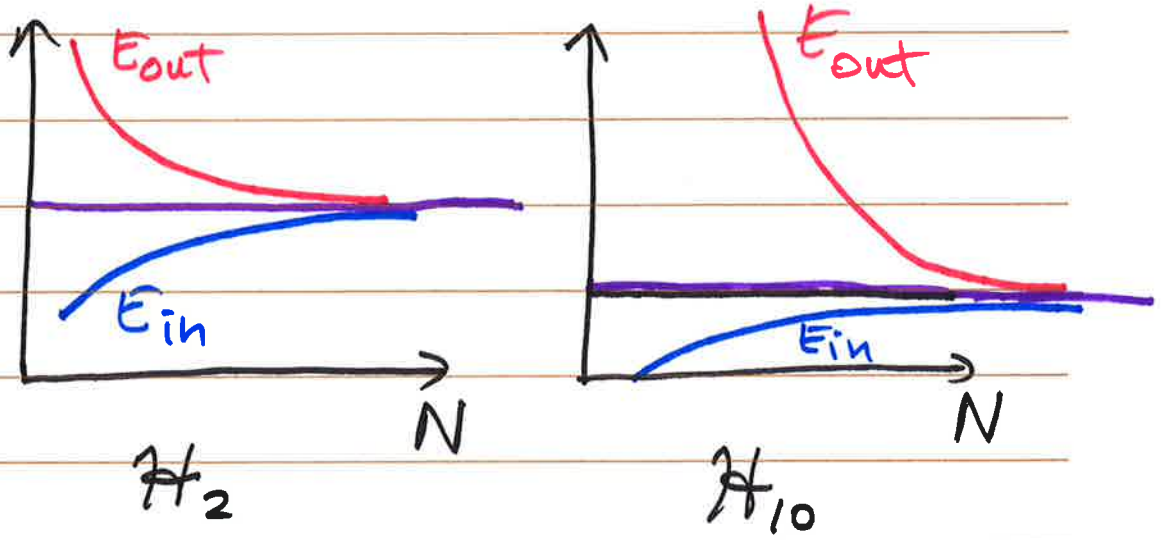
SMALL AMT. OF DATA $\rightarrow$ "DETERMINISTIC NOISE" DISTRACTS THE FIT.

SO FAR: BEST HYP. SET COMPLEXITY DEPENDS ON:

QUALITY OF DATA (NOISE) — STOCHASTIC

QUANTITY OF DATA (?) (N)

LEARNING CURVES:

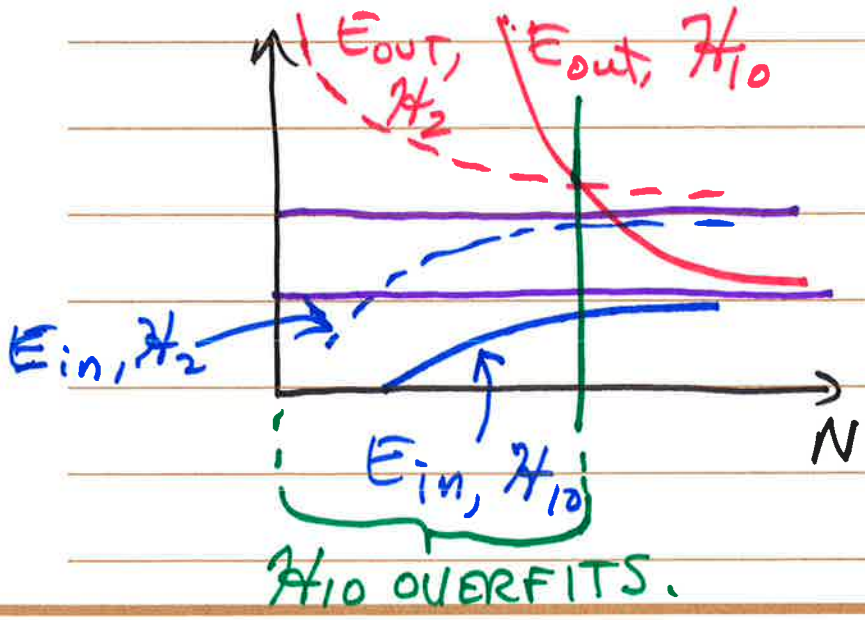


TARGET: 10^{th} ORDER POLYN., WITH NOISE.

CHECK FOR:

IF $E_{in}(h_k) \leq E_{in}(h_j)$, AND $E_{out}(h_k) > E_{out}(h_j)$

THEN WE CAN SAY h_k ~~OVERFITS~~ OVERFITS THE DATA (RELATIVE TO h_j).



⇒ BEST HYP. SET COMPLEXITY DOES DEPEND ON N (QUANTITY OF DATA).

PARAMETERS THAT AFFECT THE AMOUNT OF OVERFITTING:

N

HYP. SET COMPLEXITY

TARGET FCN. COMPLEXITY

NOISE

AML 4.1.2: MORE DETAILED EXAMPLE / SET OF EXPERIMENTS.

$f(x)$: Q_f-ORDER POLYNIMAL AS $f(x)$.

x IS UNIFORMLY DISTRIBUTED ON $[-1,+1]$.

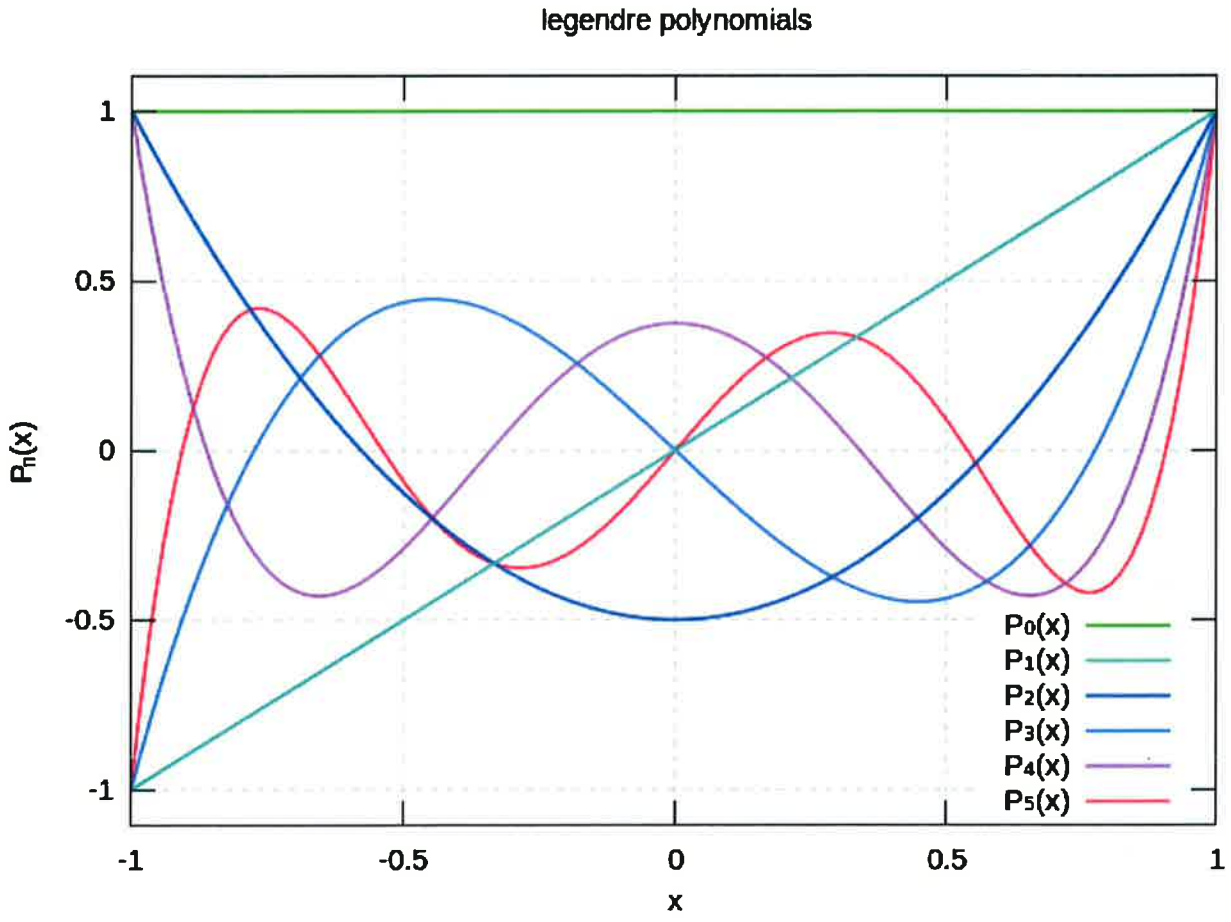
$$f(x) = \sum_{q=0}^{Q_f} a_q L_q(x)$$

L_q: q^{th} ORDER POLYNOMIAL.

a_q ARE RANDOM.

Legendre Polynomials

n	$P_n(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2-1)$
3	$\frac{1}{2}(5x^3-3x)$
4	$\frac{1}{8}(35x^4-30x^2+3)$
5	$\frac{1}{8}(63x^5-70x^3+15x)$
6	$\frac{1}{16}(231x^6-315x^4+105x^2-5)$
7	$\frac{1}{16}(429x^7-693x^5+315x^3-35x)$
8	$\frac{1}{128}(6435x^8-12012x^6+6930x^4-1260x^2+35)$
9	$\frac{1}{128}(12155x^9-25740x^7+18018x^5-4620x^3+315x)$
10	$\frac{1}{256}(46189x^{10}-109395x^8+90090x^6-30030x^4+3465x^2-63)$



Source: https://en.wikipedia.org/wiki/Legendre_polynomials

$\mathcal{D}$: N POINTS

$$y_n = f(x_n) + \sigma \epsilon_n$$

ϵ_n: I.I.D. STD. NORMAL R.V.'S.

$\Rightarrow$ FOR GIVEN N, σ, f: $E_{in}(h_{g_{10}}) \leq E_{in}(h_{g_2})$ (ALMOST ALWAYS).

USE AS OVERFIT MEASURE:

$$E_{out}(h_{g_{10}}) - E_{out}(h_{g_2})$$

AML FIG. 4.3(a): $Q_f = 20$. degree polyn. for $f(x)$.

FIG. 4.3(b): $\sigma^2 = 0.1$

→ EFFECTS OF "DETERMINISTIC NOISE".

GOOD FIGURES & EXPERIMENT FOR UNDERSTANDING AND INTERPRETING OVERFITTING. (AFTER GOING THROUGH THE PREVIOUS EXPERIMENTS AND FIGURES)