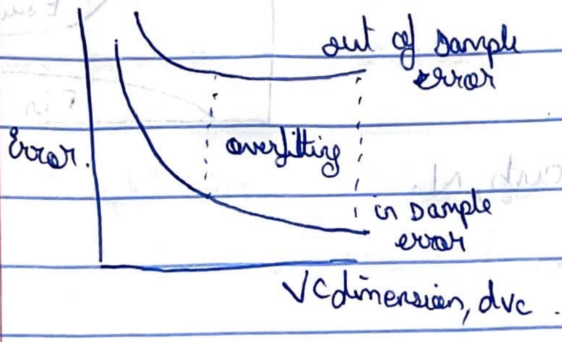
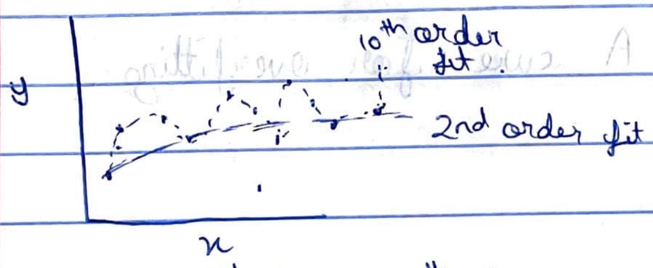
yvhEFN

Lecture -12 yvhEFN

★ ~~An~~ Overfitting:-
When lower E_{in} results in Higher E_{out}.



2^{nd} Order fit vs 10^{th} Order Polynomial.

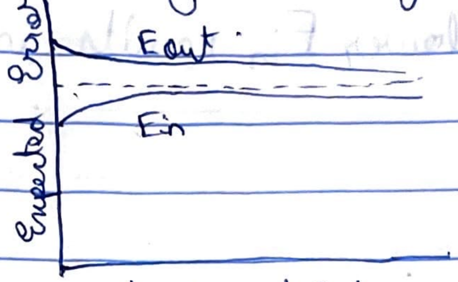


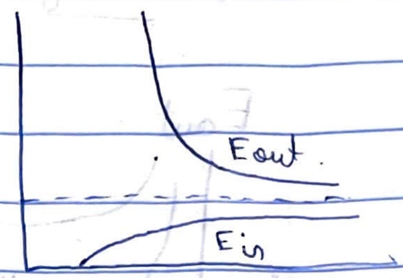
	2^{nd} order	10^{th} order
E_{in}	0.050	0.034
E_{out}.	0.127	9.00.

Sample of overfitting.

When is H_2 better than H_{10}?

Learning curves for H_2 H_{10}.

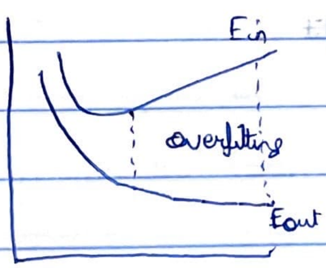




Number of Data points, N.

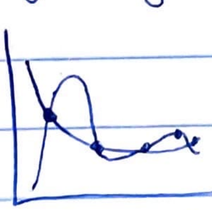
★ Regularization :-

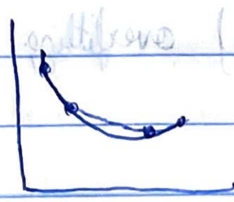
A cure for overfitting.



$E_{out}(h) \leq E_{in}(h) + \Omega(H)$ penalty complexity of H set.

Regularization : Constrain the search.





We are limiting the weight of the higher degree.

We still search for 4^{th} order for simple 4^{th} order polyno

$\text{Minimize}_{h \in H} \; E_{in}(h) + \Omega(h)$.

- Constrain the search for h with best fit to data (lowest E_{in}).

- $\Omega(h)$: a measure of the complexity of h.

- 'Weight Decay' works well for the linear model.
  Eg: Sum of squared weights.

$$h(x) = w_0 + w_1 x_1 + w_2 x_2.$$

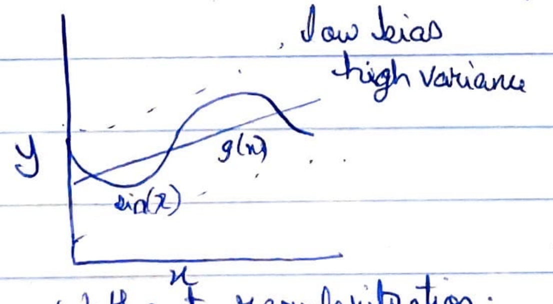
2nd order poly.

$$h(x_1, x_2) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_1 x_2 + w_5 x_2^2.$$

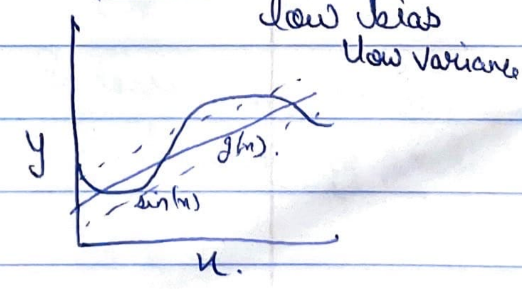
$$\Omega(h) = \sum_{j=1}^{d} w_j^2$$

★ bias :- The ability for our hypothesis to fit the data set.

★ Bias variance decomposition. Regularization



Without regularization.



Polynomials of Order Q : H_Q.

Standard Polynomial:

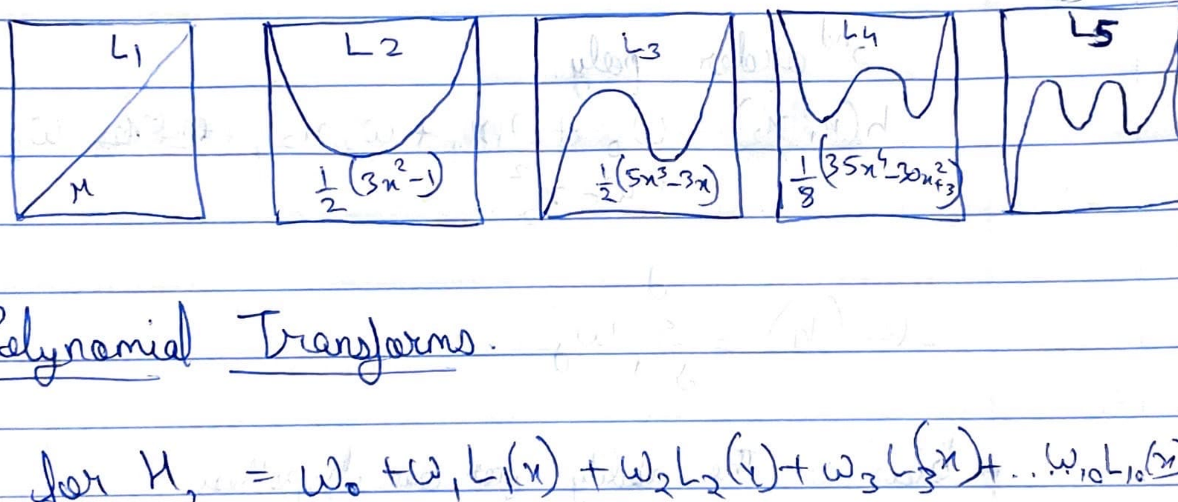
$$Z = \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^Q \end{bmatrix}$$

$$h(x) = w^T z(x)$$
$$= w_0 + w_1 x + \dots w_Q x^Q$$

Legendre Polynomial.

$$Z = \begin{bmatrix} 1 \\ L_1(x) \\ L_2(x) \\ \vdots \\ L_Q(x) \end{bmatrix}$$

$$h(x) = w^T z(x)$$
$$= w_0 + w_1 L_1(x) + \dots$$
$$w_Q L_Q(x).$$



- Polynomial Transforms.

for $H_2 = w_0 + w_1 L_1(x) + w_2 L_2(x) + w_3 L_3(x) + \dots w_{10} L_{10}(x)$

↳ set all the weights from w_3 to w_{10} as 0.

This is how we constrain the weights.

Soft order constraint: $\sum_{q=0}^{Q} w_q^2 \leq C$

set budget.

$$H_C = \{ h \mid h(x) = w_0 + w_1 L_1(x) + w_2 L_2(x) + \dots w_{10} L_{10}(x)$$
$$\text{such that } \sum_{q=0}^{10} w_q^2 \leq C \}$$

High level idea $H_C \subseteq H_{10}$

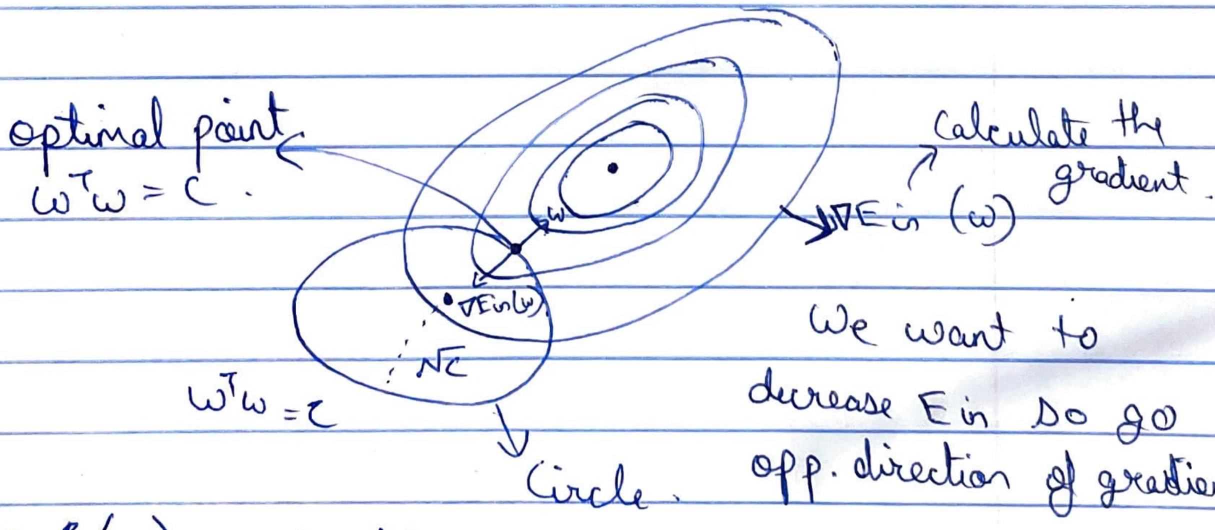
$\Rightarrow$ E_{out} closer to E_{in}.

★ Fitting the data with the soft Order Constraint.

- Want: Optimal Weights $w_{reg} \in H_C$.

$$\arg\min E_{in}(w) = \frac{1}{N}(Zw - y)^T (Zw - y)$$

subject to: $w^T w \leq C$.



We want to decrease E_{in} so go opp. direction of gradient.

$$\nabla E_{in}(w) = -2\lambda w$$

Lagrange multiplier.