SGd4af

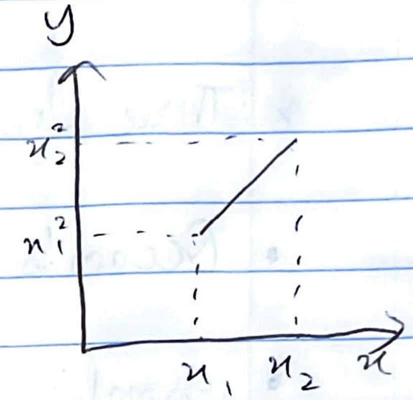
Lecture-15 — SGd4af

★ Homework 2 - Question discussion

y-value.

$$D = \{ (x_1, x_1^2), (x_2, x_2^2) \}$$

$(-1, 1)$



$$H = \{ h(x) : h(x) = w_0 + w_1 x$$
$$w_0, w_1 \in R.$$

$$w_1^2 = w_0^* + w_1^* x_1$$
$$w_2^2 = w_0^* + w_1^* x_2.$$

$$w_0^* = -x_1 x_2$$

$$w_1^* = x_1 + x_2.$$

$$g^{(D)}(x) = w_0^* + w_1^* x.$$

$$\bar{g}(x) = E_D [g^{(D)}(x)]$$

$$= E_D [-x_1 x_2 + (x_1 + x_2) x]$$

$$= E_D(x_1) = 0 \quad \text{[Mean]}$$

$$x_2 = 0$$

$$E_D[E_{out}(x)] = E_D[(\hat{y}-y)^2]$$
$$= (0-x^2)^2$$
$$= x^4$$

* Three learning principles.

- Occam's Razor: Select a Model Carefully.
  ↳ Simple model.
- Sampling Bias: Generate Data Carefully.
  ↳ Avoid.
- Data Sampling :- ~~Generate~~ Handle data Carefully.
  ↳ Avoid.

+ <u>Occam's Razor</u>:-

An explanation of the data should be made as simple as possible but not simpler.

Simpler?

- Simplicity of a Hypothesis h: $\Omega(h)$
- low order polynomial
- Hypothesis with small weights
- Easily described.

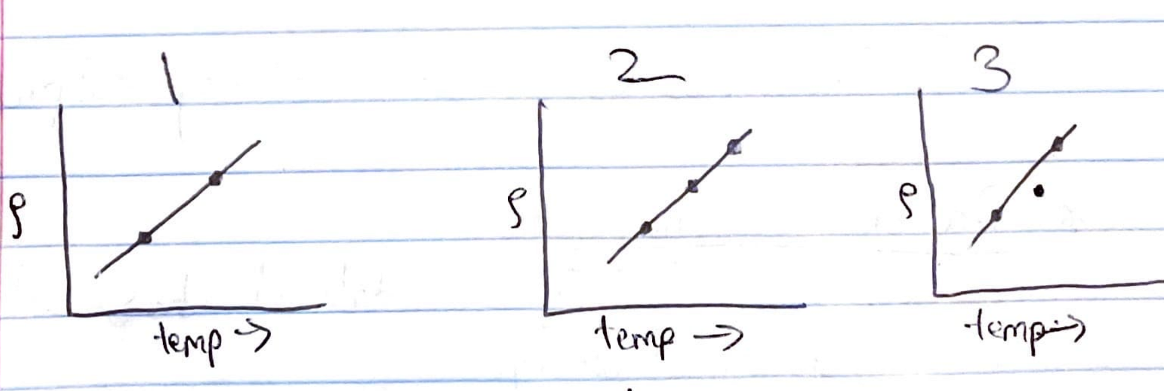
- Simplicity of a Hypothesis set H: $\Omega(H)$
- Low d_{vc}
- Small number of hypotheses
- Low entropy set.

Why is a simpler hypothesis better?

$\hookrightarrow$ More surprising when you fit the data.

Scientific Experiment

$\rho \Rightarrow$ resistivity



less data
Not Convincing

Very Convincing

Somewhat Convincing

H_2 & H_{10}.

* If data ↓ & model is complex ↑. then getting 0 Error is not convincing Enough.

* We should give the data chance to win by choosing a simpler hypothesis. Where the hypothesis want perform ~~good~~ excellent on a data. To avoid overfitting.

* <u>Sampling Bias</u>:- Randomly pick people from phone directory And Asking them whom did they vote for.
  ↳ This was a biased data collection and covered only rich people will telephone.
  ↳ This resulted in wrong output prediction

* <u>Data snooping</u>:- If a dataset has affected any step of learning.
  ↳ Here are many ways to affect data.
  ↳ Here we getting Eout should not be affected by any other criteria or external factors.

Data snooping can be subtle.

- The data looks linear, so i will use a linear model X.
- We should not torture the data to find the function which fits the data.
- Instead we should do cross-validation.

- Literature survey of things others have tried. I will improve upon it.
- Input normalization: normalize all data together. and set aside a portion as test set.

*

- * VC analysis :- Approximation VS Generalization.
- * Linear model for 3 learning problems. (credit analysis)
- * Mechanics of non-linear feature transforms.
- * Use feature transforms Responsibly.
- * Overfitting.
- * 2^{nd} order vs 10^{th} order Polynomial.
- * Overfitting Measure (When to choose simpler vs complex) models.
- * Regularization (To combact overfitting).
  - ↳ Controlled by term λ (We choose the value of λ)

Choose λ which is just right.

$\lambda \downarrow \Rightarrow$ Overfitting $\qquad \lambda \uparrow \Rightarrow$ Underfitting.

* three learning principles:

* **Mid Term 1**: Tuesday 10/17.

Open book. LFD chapters 1,2,3,4,5.

5-6 Reasoning / essay Questions.

5-6 or more Multiple choice Questions.

↳ If multiple correct then mark all correct.

*