

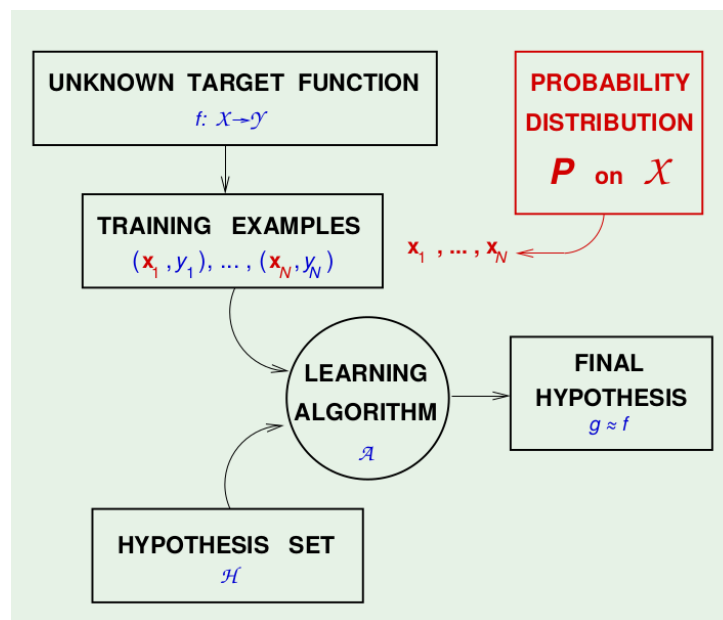
# MAC0460/5832 – Introdução ao Aprendizado de Máquina

DCC / IME-USP — Primeiro semestre de 2020

## Lista de exercícios 1

Esta lista cobre o conteúdo sobre *VC theory* (especificamente, as *Lectures* 02, 05, 06 e 07). Responda ou comente as questões e entregue uma versão escaneada ou fotografada, ou ainda um pdf (caso decida digitar as respostas). Note que não há, necessariamente, o certo ou não-certo para as respostas. A lista pretende avaliar qual compreensão sobre o assunto cada um alcançou e também pretende servir como guia para auto-avaliação e estudos adicionais individual.

1. Comente sobre o diagrama abaixo. O que o diagrama como um todo ilustra e o que cada componente representa?



2. O que é  $E_{in}$  e  $E_{out}$ ? Dê um exemplo concreto.
3. Por que apenas minimizar  $E_{in}$  não é suficiente? Em alguma situação pode ser suficiente?
4. Quando consideramos a formulação teórica de aprendizado de máquina, uma das possibilidades é investigar o valor  $|E_{in} - E_{out}|$ . O que esse valor expressa e por que nos interessa investigar ele?

5. O prof. Abu-Mostafa menciona recorrentemente o termo hipótese. A o que ele se refere quando fala em hipótese?
6. A desigualdade de Hoeffding, no contexto de aprendizado de máquina, com respeito a uma certa hipótese  $h$ , é dada por:

$$P\left(|E_{in}(h) - E_{out}(h)| > \epsilon\right) \leq 2e^{-2\epsilon^2 N}$$

Explique o significado dessa desigualdade.

7. A desigualdade de Hoeffding, no contexto de aprendizado de máquina, com respeito a uma situação em que o espaço de hipóteses consiste de  $M$  exemplos é dada por:

$$P\left(|E_{in}(g) - E_{out}(g)| > \epsilon\right) \leq 2Me^{-2\epsilon^2 N}$$

Comente o significado da diferença entre essa desigualdade e a do item anterior.

8. O *bound*  $2Me^{-2\epsilon^2 N}$  no item anterior foi obtido aplicando-se o *union-bound*. O que é *union-bound*?
9. Que elementos são relevantes na definição de dicotomias? O que são dicotomias geradas por um espaço de hipóteses  $\mathcal{H}$ ?
10. O que é *growth-function*?
11. No contexto sendo considerado, qual o interesse em dicotomias e *growth-function*?
12. Qual é o interesse em se provar que o *growth-function* é polinomial?
13. Por que não podemos simplesmente trocar o  $M$  em  $2Me^{-2\epsilon^2 N}$  pelo *growth-function*  $m_{\mathcal{H}}(N)$ ?
14. Relacione dicotomias e *VC dimension*.
15. Qual é o *VC dimension* do perceptron? Como é feita a demonstração?
16. Dissemos que o *VC dimension* relaciona-se com a expressividade do espaço de hipóteses. Comente sobre isso.

17. Ao final, o bound inicial  $2Me^{-2\epsilon^2N}$  acaba sendo substituído por um *bound* que depende do *VC dimension*. Qual é esse novo *bound* ? E o que podemos dizer sobre esse novo *bound*, em termos do *VC dimension*?
18. De acordo com a desigualdade VC, como podemos calcular o número de amostras necessárias para se garantir uma certa precisão  $\epsilon$ , com probabilidade 0.9, por exemplo ?
19. De acordo com a desigualdade VC, dado uma certa quantidade  $N$  de exemplos de treinamento, o que podemos dizer sobre  $|E_{in}(h) - E_{out}(h)|$  ?
20. Por que garantir apenas  $|E_{in}(h) - E_{out}(h)| < \epsilon$  não é suficiente ?
21. Na sua opinião, o *VC bound* poderia ser melhorado? O que poderia ser explorado para se obter um *bound* teórico menor do que o *VC bound*?
22. As *Lectures* 02, 05, 06 e 07 cobrem o *VC theory* (embora em contexto restrito à classificação binária). Escreva aqui, com suas palavras e de forma sucinta, como você explicaria o *VC theory* para um colega interessado em aspectos teóricos de aprendizado de máquina.