

## MAC0460 - Introdução ao aprendizado de máquina

### Lista 2

Arthur Font Gouveia - 12036152

**1.** O diagrama como um todo ilustra o processo de aprendizado de máquina. No qual cada componente contribui para gerar uma função que aproxima uma função alvo. Podemos descrever cada componente representa da seguinte maneira:

- Uma função alvo desconhecida  $f : X \rightarrow Y$  que gera todos os exemplos de treinamento, onde  $X$  é o espaço de entrada (conjunto de todas as entradas possíveis) e  $Y$  é o espaço de saída (conjunto de todas as saídas possíveis)

- Um conjunto de exemplos de treinamento. Esse conjunto de dados de exemplos de entrada e saída  $(x_1, y_1), \dots, (x_N, y_N)$ , onde  $y_n = f(x_n)$  para  $n = 1, \dots, N$ .

- Um espaço de hipóteses  $H$  que representa uma abstração de todas as possíveis funções que nosso algoritmo pode gerar. Chamamos de hipótese um elemento  $h \in H$ .

- Um algoritmo de aprendizado  $A$  que, a partir dos exemplos de treinamento, gera uma função que aproxima  $f$ . A escolha do algoritmo a ser utilizado se dá conforme a necessidade.

- Uma hipótese final  $g \approx f$ . Essa hipótese é o resultado de nosso algoritmo de aprendizado, a partir dela, podemos obter novos resultados de acordo com novas observações do mundo, e esses resultados se aproximam ao esperado caso fossem calculados com a função  $f$ .

**2.**  $E_{in}$  é a taxa de erro que acontece na amostra, ou seja, a porcentagem de pontos selecionados erroneamente dentro do espaço amostral.  $E_{out}$  é a taxa de erro que acontece fora da amostra, ou seja, a porcentagem de pontos selecionados erroneamente fora do espaço amostral.

**3.** Esse valor nos mostra a diferença de performance da nossa hipótese entre os dados de treinamento e os dados de mundo real. Nos interessa investigá-lo pois através do mesmo é possível obter a generalização do erro, e assim evitar que o algoritmo sofra com o superajuste de parâmetros.

**4.** Essa desigualdade indica que quanto maior o tamanho da amostra e maior a margem de erro, menor a probabilidade de ultrapassar essa margem de erro. Portanto, essa desigualdade mede a qualidade de generalização de uma hipótese.

5. A diferença entre essa desigualdade e a do item anterior é que essa desigualdade possui o parâmetro  $M$ , que representa o número de hipóteses de  $H$ . Portanto, desigualdade mede a qualidade de generalização de todo um espaço de hipóteses.
6. Union bound é uma aproximação da probabilidade  $P$  da disjunção de vários eventos.
7. Dicotomia é o conjunto de valores gerados por uma hipótese. *Growth-function* é uma função que conta o número máximo de dicotomias que podem ser geradas a partir de um espaço de hipóteses  $H$  usando  $N$  observações. Se nenhum conjunto de dados for quebrado por  $H$ , então  $k$  é um *break-point* para  $H$ . A relação entre eles é que podemos utilizar o conceito de *break-point* para limitar a *Growth-function*.
8. O interesse em realizar essa troca é que, se a função *Growth-function* for polinomial, o bound obtido pode ser bem pequeno ao escolher o  $N$  adequado.
9. Quanto maior for a quantidade de parâmetros que um modelo possui, mais diverso será o seu conjunto de hipóteses, o que reflete em um valor maior da *growth function*. A *VC dimension* mede a efetividade desses parâmetros que permite o modelo expressar um conjunto diversificado de hipóteses.
10. O VC bound define um limite para o valor que  $E_{out}$  pode assumir.
12. A desigualdade não é suficiente pois caso os valores de  $E_{in}$  e  $E_{out}$  sejam altos e próximos, o valor de  $e$  seria pequeno, mesmo que a hipótese não seja boa para os conjuntos de treinamento e teste.
13. Ambos estimam o  $E_{out}$  (ou erro de generalização). O *VC analysis* limita o erro através do  $E_{in}$  e o tamanho da amostra, obtendo melhor performance ao estimar o  $E_{out}$  em grandes amostras. Já o *Bias-variance analysis* utiliza a amostra para ajustar os parâmetros, logo apresenta pior performance ao estimar o  $E_{out}$ .
14. Os conteúdos cobertos nas *lectures* ajudam a compreender como funciona o processo de aprendizado de máquina, e a relação com a amostra e os parâmetros do algoritmo de aprendizado estabelecido.