"Uma Breve Introdução à Teoria da Aprendizagem Computacional"

12 de março de 2019; rev. 13 de março de 2019

Fred Guth

Preâmbulo

Complexidade Computacional, que aqui na UnB vemos no curso de Projeto e Complexidade de Algoritmos é um dos alicerces de qualquer curso de Ciência da Computação. A ideia desse campo de estudo é examinar e classificar problemas de acordo com sua dificuldade de solução (P vs NP, por exemplo). Ao fazer esse curso no semestre passado e tendo como área de interesse aprendizado de máquinas, me perguntei: Será que não seria possível trazer esses conceitos de complexidade de algoritmos para aprendizado de máquinas, analisando qual seria a quantidade de amostras necessárias para um algoritmo aprender uma tarefa?

A resposta é sim. Essa ideia não é, obviamente, original. Em 1984, Leslie Valiant publicou o modelo de aprendizado "Provavelmente Aproximadamente Correto"(PAC, Probably Approximately Correct), justamente com o objetivo de incitar pesquisadores que estudavam complexidade de algoritmos a pensar problemas de aprendizado. Ele introduziu a ideia de problemas de aprendizado que são apreensíveis em tempo polinomial, PAC apreensíveis, em analogia com a classe dos problemas P. Pode-se dizer que foi bem sucedido: diversos pesquisadores estenderam ou propuseram novas teorias, o que originou o ramo de estudo chamado de Teoria de Aprendizado Computacional, que tem até a sua própria conferência anual, COLT (Conference on Learning Theory).

Nesta apresentação meu objetivo é apresentar o modelo PAC, contextualizar a Teoria do Aprendizado na minha linha de pesquisa e me colocar a disposição de alunos e professores interessados para discutir mais este assunto.

O que é Apreensível?

O que é apreensível? Essa é uma questão muito anterior à Ciência da Computação. No século 18, o filósofo escocês David Hume se perguntou se é possível gerar conhecimento a partir da indução (O problema da Indução). O que é justamente a base do aprendizado supervisionado, a área da Inteligência Artificial mais pesquisada no momento.

Para Hume, não há justificativa lógica para generalizações baseadas em indução. Por exemplo: "Sempre que no passado comi pão, ele me alimentou. O futuro será como o passado. Portanto, da próxima vez que comer pão ele me alimentará.", não tem justificativa. Para ele nós vemos causalidade onde há apenas a conjunção constante entre dois acontecimentos distintos. Mas se isso é verdade, é possível obter conhecimento através da indução? E se não é possível, o aprendizado indutivo supervisionado não tem justificativa? Pior, o método científico não pode ser logicamente justificado?

A abordagem de Valiant traz uma resposta ao mesmo tempo prática e formal para esse problema. Não é possível dizer que "esse pão que ainda não comi vai me alimentar", sem comê-lo. Entretanto, é possível dizer, com certo nível de confiança que a hipótese "vai me alimentar"está aproximadamente correta. Em outras palavras, embora seja possível que o pão não me alimente (vai que está estragado?), na maioria das vezes ele me alimenta e é possível mensurar a probabilidade desta hipótese estar aproximadamente correta.

Confuso? Vamos colocar mais matemática nessa discussão para deixar tudo mais claro. Só que antes, algumas definições.

Aprendizado-PAC: Definições

Genericamente, podemos pensar em um algoritmo como uma função que transforma uma instância do espaço de entrada, $\mathcal{X}$, em uma instância do espaço de soluções da tarefa que queremos realizar, $\mathcal{Y}$. Em geral, essa função é um conjunto de regras programadas.

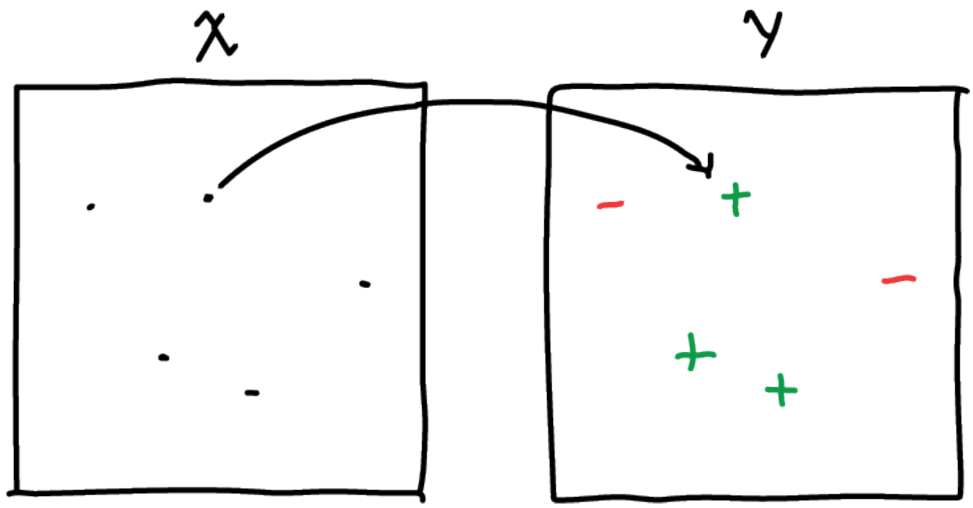

Figura 1: Um algoritmo genérico: $\mathcal{X} \to \mathcal{Y}$.

No contexto de aprendizado de máquina, não sabemos expressar essa função alvo, portanto teremos que inferi-la, aprendê-la. Chamaremos essa função ideal de *conceito*. É o conceito que queremos aprender.

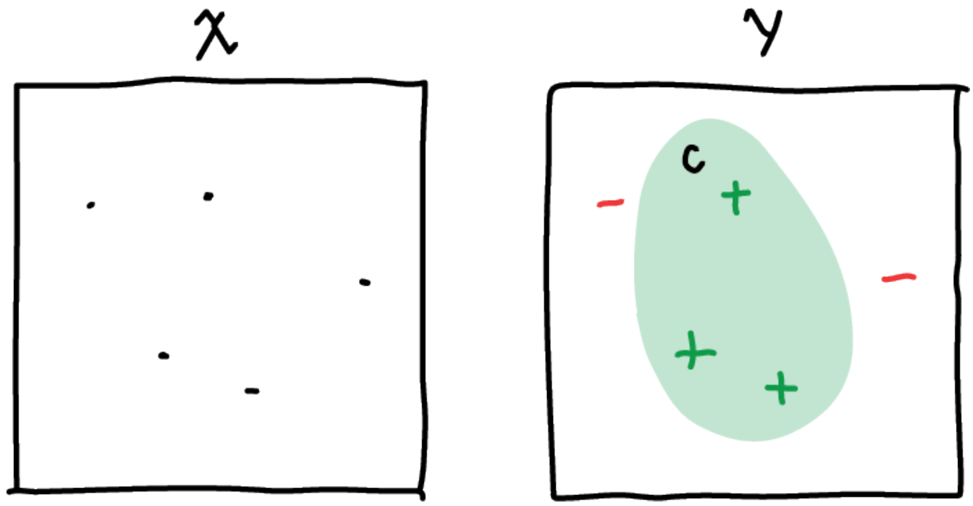

Figura 2: Um conceito c.

A função realmente obtida, a regra, a heurística, chamamos de *hipótese*. Uma hipótese pertence a um espaço de Hipóteses: $h \in \mathcal{H}$.

References

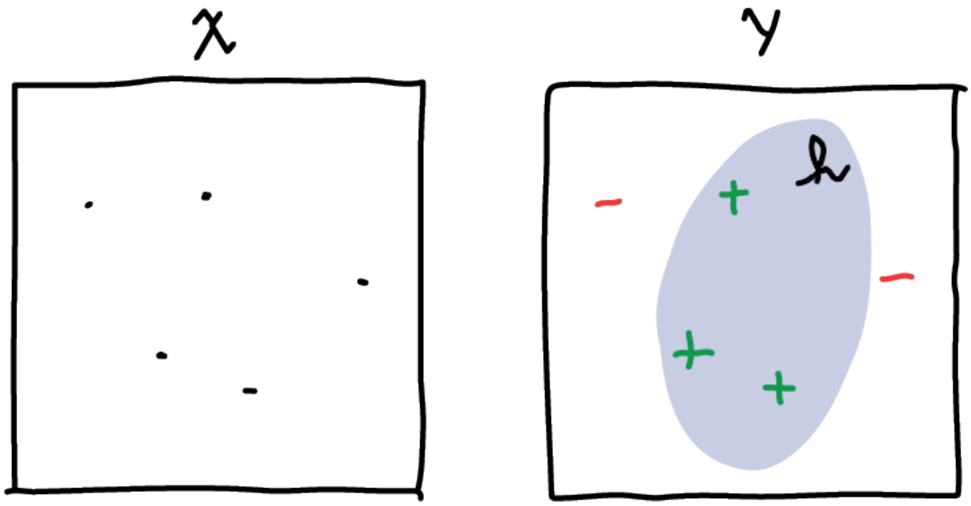

Figura 3: Uma hipótese $h \in \mathcal{H}$.