

Universidade do Estado do Amazonas

Escola Superior de Tecnologia

Data: 22 de maio de 2018

Disciplina: Tópicos Especiais *Machine Learning* - 2018.1

Professor(a): Elloá B. Guedes

Alunos: Iêsa B. Lobato, Lucas P. Reis, Victor S. Lopes

Curso: Engenharia da Computação

MINI TESTE #4

Aprendizagem de Máquina (ou *Machine Learning*) é um subramo da inteligência artificial em expansão devido a alta necessidade das empresas de obter resultados a partir da grande quantidade de dados que são gerados. Esses modelos preditivos evoluíram de tal forma que aparentam estar presentes em todos os lugares devido a sua incrível versatilidade em várias aplicações. Embora seja possível atuar em diferentes tarefas, ainda existem problemas e contextos em que técnicas e métodos da Aprendizagem de Máquina não se aplicam.

Uma exigência para começar a utilizar as ferramentas de *Machine Learning* é possuir uma grande massa de dados de boa qualidade. Ou seja, é preciso ter certeza que os dados obtidos estão estruturados corretamente e livres de anomalias [5]. Além disso, existem casos em que é necessário reestruturar os dados de acordo com a ferramenta utilizada. Embora o método consiga automatizar o processo de análise, o trabalho de filtrá-los ainda é feito por profissionais. Essa atividade é conhecida como *data cleaning*, e consiste em corrigir o conjunto de dados de forma que este se torne coerente ao problema.

Grande parte dos modelos preditivos atuais são treinadas através de aprendizagem supervisionada, e para este treinamento ocorrer corretamente, os dados presentes no *dataset* necessitam, primeiramente, ser rotulados. Esta tarefa atualmente está sendo feita completamente por humanos, pois ainda não existe um modelo eficiente capaz de verificar um dado e classificá-lo corretamente sem possuir um conhecimento prévio[1]. Em alguns *datasets* o rótulo dos dados são cruciais para a predição como, por exemplo, um *dataset* composto por imagens para classificá-las em alguma categoria. A rotulação é um tópico relevante para a predição mas que não pode ser resolvido por aprendizagem de máquina. Normalmente as grandes empresas ao redor do mundo estão contratando pessoas, denominadas de *data labelers*, para rotular seus dados, algo que está sendo bastante recorrente nos dias de hoje [1].

Outro ponto importante sobre pré-processamento dos dados é que, dentre os dados disponíveis no *dataset*, o modelo criado não é capaz de identificar quais serão relevantes ou não. O modelo precisa ser treinado para ser capaz de interpretar seus próprios dados para então, posteriormente, afirmar quais dados serão importantes. Isto nos leva a outro assunto importante: atualmente os modelos preditivos não são capazes de interpretar seus dados, ou um modelo interpretar os resultados de outro. Esses dois contextos explicados são realizados por humanos.

Realizar um pré-processamento nos dados é um passo extremamente importante para treinar um modelo preditivo, mas se estiver sendo desenvolvido uma aplicação ou resolvendo um problema onde a quantidade de dados utilizados for muito baixa, as decisões do modelo não serão satisfatórias, visto que a aprendizagem de máquina é designada para

trabalhar com uma quantidade muito grande de dados. Além disso se o *dataset* apresentar muitos dados ruidosos e não realizar um pré-processamento para removê-los, o modelo aprenderá características irrelevantes para o problema em questão [2].

Generalizando a ideia vista anteriormente, podemos afirmar que os modelos preditivos não são capazes de chegar às conclusões de determinados contextos ou não são capazes de ter *insights* sobre resultados. Além do exemplo já mencionado, também pode-se citar que através de aprendizagem de máquina é possível criar textos, mas não é possível criar textos, artigos e/ou livros que foram desenvolvidos a partir das experiências do escritor e de suas bagagens de conhecimento e inteligência [3]. Com aprendizagem de máquina já é possível realizar leitura labial e traduzir falas ou textos, mas ainda não é possível obter o contexto ou situação de algum vídeo dado como entrada [3].

Contextualizar uma situação é entendê-la. Não apenas identificar seus componentes, mas também o significado do conjunto formado por eles. O empecilho se concentra no seguinte fato: máquinas não compreendem a linguagem humana, seriam necessárias muitas regras para explicar cada palavra e complexidade gramatical. Segundo o professor Hubert Dreyfus, no livro *What Computers Can't Do* (1972), o problema é que muitas coisas requerem um tipo de inteligência instintiva que não pode ser capturada com regras [4].

O *hardware* do computador onde o modelo preditivo aprenderá é um fator que pode ser decisivo em uma aplicação, pois caso o problema seja muito complexo o aprendizado poderá se tornar altamente custoso, havendo a necessidade de realizar gastos para aumentar a capacidade de processamento. Às vezes, pode-se até mesmo causar a inviabilidade do projeto de utilizar aprendizagem de máquina [2]. Por outro lado, caso o problema seja bem simples de resolver ou seu comportamento já é conhecido, podendo obter seus valores exatos através de funções, não é necessário ou viável a utilização de aprendizagem de máquina [2].

De forma geral, quando se trata de Aprendizagem de Máquina, aplica-se o conceito do Princípio de Pareto. Tal princípio afirma que para muitos fenômenos, 80% das consequências advêm de 20% das causas. Apesar de uma ferramenta poderosa, a Aprendizagem de Máquina ainda assim é apenas um co-piloto que demanda a ação de um ser humano para analisar seus resultados. Assim como a Aprendizagem de Máquina pode realizar tarefas não viáveis ao ser humano, como organizar anos de dados em segundos, humanos podem realizar tarefas que a máquina não é capaz, como analisar se dado padrão gerado faz sentido ou é apenas uma anomalia. Portanto, mesmo tendo em vista o crescente avanço nas tecnologias e ferramentas associadas a Aprendizagem de Máquina, as quais facilitam 80% da carga de trabalho, a presença de Cientistas de Dados é indispensável para análises mais profundas e tomadas de decisões precisas.

[1] QUATERFLY, McKinsey. **What AI can and can't do(yet) for your business**, *McKinsey*. Disponível em: <<https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/what-ai-can-and-cant-do-yet-for-your-business>>.

[2] ZACHARIASZ, Kataryzna. **When to Use Machine Learning - Does Your App Really Need ML?**, *Netguru*. Disponível em: <<https://www.netguru.co/blog/when-to-use-machine-learning-does-your-app-really-need-ml>>.

[3] FRICK, Walter. **Why AI Can't Write This Article (Yet)**, *HBR*. Disponível em:

<<https://hbr.org/2017/07/why-ai-cant-write-this-article-yet>>.

[4] KNIGHT, Will. **AI's Language Problem**. Disponível em: <<https://www.technologyreview.com/s/602094/ais-language-problem/>>.

[5] BOYD, Mark. **What Machine Learning Can and Can't Do**. Disponível em: <<https://thenewstack.io/what-machine-learning-can-and-cant-do/>>.