

## Materiais:

**Q: Onde encontro os materiais de apoio e os arquivos notebooks para execução dos exercícios?**

R: O material de apoio consta no ícone de folha a4 ao lado do título da disciplina.

## Ferramentas:

Não houve comentários sobre ferramentas até a data de geração deste FAQ.

## Conceitos e Exercícios:

**Q: O aprendizado auto-supervisionado é o mesmo que aprendizado por reforço? Se não, onde se encaixa o aprendizado por reforço e no que se difere do auto-supervisionado?**

R: Diferenciar o aprendizado auto-supervisionado do aprendizado por reforço é mais simples, mas a linha entre auto-supervisionado e não supervisionado é bastante tênue. Na primeira aula do professor Thomas (Aula 3), ele dá uma explicação sobre essas diferenças.

De forma geral, o **aprendizado auto-supervisionado** não utiliza rótulos, mas tem algum tipo de mecanismo para "prover supervisão". Por exemplo, nos casos de processamento de linguagem natural como BERT, parte das palavras das frases é removida e o algoritmo é treinado para prever as palavras. Nesse caso, não existe um "rótulo", mas ao remover as palavras e tentar prevê-las, podemos entender que é um "método não supervisionado treinado de forma supervisionada".

O **aprendizado por reforço** é bastante diferente onde a principal diferença está no mecanismo de aprendizado - é um paradigma que busca aprender por tentativa e erro, e utiliza um mecanismo de feedback para guiar seu aprendizado. O objetivo é aprender uma "política", que mapeia um "estado" para uma sequência de ações que maximizam tal recompensa ao longo do tempo. Em outras palavras, modelos de aprendizado de máquina são treinados para tomar uma sequência de decisões. O agente aprende a atingir uma meta em um ambiente incerto e potencialmente complexo. No aprendizado por reforço, o sistema de inteligência artificial enfrenta uma situação. O computador utiliza tentativa e erro para encontrar uma solução para o problema. Para que a máquina faça o que o programador deseja, a inteligência artificial recebe recompensas ou penalidades pelas ações que executa. Seu objetivo é maximizar a recompensa total. Um ponto interessante é que, embora o Cientista de Dados defina a política de recompensa - isto é, as regras do jogo - ele não dá ao modelo nenhuma dica ou sugestão de como resolver o jogo. Cabe ao modelo descobrir como executar a tarefa para maximizar a recompensa, começando com testes totalmente aleatórios e terminando com táticas sofisticadas.

Também na Aula 3, o professor Thomas traz alguns exemplos desse tipo de aprendizado. Um exemplo prático de aprendizado por reforço seria aprender a sequência de movimentos que um braço robótico deve realizar para conseguir pegar um objeto, ou a sequência de movimentos que um jogador de um determinado jogo deve realizar para completar uma fase.

*\*FAQ gerado com base em comentários até o dia 02/03/2022.*