

<b>Curso</b>	Ciência da Computação	<b>Ano:</b>	<b>Data:</b>
<b>Disciplina</b>	Computação e Tecnologia		
<b>Professor</b>	Prof. Eduardo Pena		
<b>Laboratório sobre Machine learning</b>			
Orientações gerais:			

Responda as seguintes perguntas.

1. Como você definiria Machine Learning (ML)?
2. Você pode citar três tipos de problemas onde ML se destaca?
3. O que é um conjunto de treinamento rotulado?
4. Que tipo de algoritmo de aprendizado de máquina você usaria para permitir que um robô ande por vários terrenos desconhecidos?
5. Que tipo de algoritmo você usaria para segmentar os clientes de um e-commerce em vários grupos?
6. Você enquadraria o problema da detecção de spam como um problema de aprendizado supervisionado ou problema de aprendizado não supervisionado?
7. Utilizando as bases de dados:

- <http://archive.ics.uci.edu/ml/index.php>
- <https://www.kaggle.com/datasets>
- <https://registry.opendata.aws/>

Encontre um dataset (conjunto de dados) que ache interessante para preencher a seguinte definição:

A computer program is said to learn from experience **E** with respect to some task **T** and some performance measure **P**, if its performance on **T**, as measured by **P**, improves with experience **E**.

Você deve apontar o dataset, e apontar pelo menos um artigo científico onde aquele dataset foi usado. Descreva pelo menos uma das tarefas (task **T**) em que o dataset tem sido utilizado. Descreva pelo menos uma métrica (performance measure **P**) utilizada com aquele dataset, e como os dados estão organizados (experience **E**).

Veja o último slide disponível no moodle para um exemplo.

8. Considere o notebook (código Python) que acompanha a atividade. Ele inclui apenas um classificador, baseado em **decision trees** ([https://en.wikipedia.org/wiki/Decision\\_tree](https://en.wikipedia.org/wiki/Decision_tree)). Pesquise no site do scikit-learn ([https://scikit-learn.org/stable/supervised\\_learning.html#supervised-learning](https://scikit-learn.org/stable/supervised_learning.html#supervised-learning)) um algoritmo adicional e inclua-o

no código. Execute o novo algoritmo e calcule a acurácia de suas classificações (siga o modelo passado). A acurácia do novo algoritmo é melhor do que aquela da Decision Tree? Utilize o esquema de visualização disponível no arquivo para visualizar graficamente quais dígitos seu algoritmo tem mais dificuldade em classificar corretamente.