

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS CORNÉLIO PROCÓPIO
PRÓ-REITORA DE GRADUAÇÃO E EDUCAÇÃO PROFISSIONAL
ENGENHARIA DA COMPUTAÇÃO

PRÁTICAS DE APRENDIZADO DE MÁQUINA

**RELATÓRIO DE ATIVIDADES
ESTÁGIO À DOCÊNCIA**

LUCAS M. VALERIO, PRISCILA T. M. SAITO

CORNÉLIO PROCÓPIO

2019

SUMÁRIO

1	INTRODUÇÃO	3
1.1	PLANO DE PROJETO	3
1.2	CONJUNTOS DE DADOS, CÓDIGOS E MATERIAIS	4
1.3	REGRAS GERAIS E CRITÉRIOS DE AVALIAÇÃO	4
2	PRÁTICA 1	6
2.1	ENUNCIADO	6
2.2	USO DA FERRAMENTA	6
3	PRÁTICA 2	7
3.1	ENUNCIADO	7
4	PRÁTICA 3	8
4.1	ENUNCIADO - SUGESTÕES	8
5	PRÁTICA 4	9
5.1	ENUNCIADO	9
6	PRÁTICA 5	10
6.1	ENUNCIADO	10

1 INTRODUÇÃO

Este projeto tem como objetivo introduzir atividades práticas relacionadas a abordagens de aprendizado de máquina.

Foram desenvolvidos códigos de exemplos para que os alunos possam se familiarizar com as abordagens e executar experimentos. Os códigos foram desenvolvidos na linguagem Python, e estão alocados na plataforma Google Colab para facilidade de execução.

1.1 PLANO DE PROJETO

- **Prática 1** - Extração de Características: nesta prática o aluno deve analisar descritores tradicionais existentes na literatura. Após a seleção dos descritores a serem analisados, deve realizar a descrição dos conjuntos de dados selecionados, gerando os arquivos de características que serão utilizados nas próximas práticas;
- **Prática 2** - Aprendizado Supervisionado: nesta prática o aluno deve analisar o desempenho de classificadores supervisionados tradicionais. A partir dos arquivos de entrada obtidos na prática 1 devem ser realizadas as classificações, de forma a realizar a análise dos resultados por meio de diferentes métricas de avaliações;
- **Prática 3** - Aprendizado Não Supervisionado: nesta prática o aluno deve explorar técnicas de agrupamento de dados (e.g. k -means) e de redução de dimensionalidade (e.g. PCA), a partir dos arquivos de entrada obtidos na prática 1;
- **Prática 4** - Aprendizado Semi-Supervisionado: nesta prática o aluno deve avaliar a propagação de rótulos realizada por meio das técnicas de aprendizado semi-supervisionado;
- **Prática 5** - Aprendizado Ativo: nesta prática o aluno deve explorar estratégias para seleção de amostras mais informativas a serem consideradas no processo de aprendizado iterativo do classificador.

1.2 CONJUNTOS DE DADOS, CÓDIGOS E MATERIAIS

Para desenvolvimento da Prática 1 serão obtidos os arquivos de entrada necessários nas demais práticas. Para tanto, a partir da seleção de algumas bases de imagens disponíveis devem ser extraídas características por meio de diferentes descritores de imagem. Algumas bases de imagens previamente organizadas e uma ferramenta que poderá ser utilizada para a descrição estão disponíveis neste [link](#). No entanto, sugere-se também a utilização de outras bases de imagens, bem como outras ferramentas para extração das características.

A maioria de exemplos de soluções para as práticas foram realizados utilizando Python e a plataforma Google Colab. Sendo assim são apresentados links úteis explicando o uso de ferramentas, métodos e particularidades da linguagem:

- Tutorial para utilização do Google Colab. Nesse caso, foi considerada a prática 2 como exemplo (disponível no [link1](#) e [link2](#)).
- Curso completo de Python realizado pelo IME-USP ([link](#)).
- Material sobre a biblioteca Pandas em Python de análise e estrutura de dados ([link](#)).
- Material sobre o agrupamento por meio do algoritmo *k*-means em Python, utilizando a biblioteca Scikit-Learn ([link](#)).
- Material sobre a biblioteca Scikit-Learn de aprendizado de máquina em Python ([link](#)).
- Documentação da biblioteca Scikit-Learn ([link](#)).

1.3 REGRAS GERAIS E CRITÉRIOS DE AVALIAÇÃO

Os alunos devem se agrupar em equipes para desenvolvimento das atividades práticas propostas em sala de aula. Descrições detalhadas das atividades a serem realizadas em cada prática serão apresentadas.

As equipes serão avaliadas por meio de entrega de relatório final. Cada aluno também será avaliado individualmente, durante participação e realização das práticas em sala de aula, bem como arguição individual ao final do desenvolvimento em data a ser definida. O relatório final consiste em um texto (máximo de 10 páginas), contendo:

- Título do trabalho e nome dos alunos/registros acadêmicos;

- Introdução, motivação e objetivos;
- Fundamentação teórica dos temas abordados em cada prática (explicação sucinta de técnicas da literatura relacionadas e/ou utilizadas. Incluir também resumo de pelo menos um artigo recente da literatura (2015+) envolvendo o tema de alguma das práticas e/ou técnicas de *Deep Learning*);
- Metodologia Proposta (propor e apresentar a descrição detalhada de uma estratégia para seleção de amostras mais informativas, conforme estabelecido na prática 5);
- Avaliação experimental (descrição das técnicas implementadas e/ou dos experimentos realizados, incluindo descrição dos conjuntos de dados selecionados e cenários considerados);
- Resultados/discussões (análise e discussão dos resultados obtidos);
- Conclusões e trabalhos futuros (apresentar as conclusões/considerações finais).

Sugestões:

- O relatório pode ser escrito em template utilizado em publicações IEEE, em latex ou .doc ([link template](#)). Entregar arquivos fontes e arquivo .pdf.
- Selecionar o artigo de alguma destas fontes: [IEEE Xplore Digital Library](#), [ACM Digital Library](#), [Google Acadêmico](#).

2 PRÁTICA 1

Nesta prática poderá ser utilizada uma ferramenta em JAVA para extração de características de cada imagem dos conjuntos de dados selecionados. Ao final, pretende-se obter os arquivos em formato .arff, os quais serão utilizados nas próximas práticas.

2.1 ENUNCIADO

Selecione pelo menos 2 conjuntos de dados e realize a extração de características por meio de descritores tradicionais da literatura.

Observações:

- Selecionar os conjuntos de dados considerando diferentes cenários (considerando quantidades de amostras, classes e complexidade);
- Selecionar os extratores de imagens considerando diferentes categorias e quantidades de características a serem extraídas;
- Armazenar apenas os arquivos no formato .arff, os quais serão gerados pela ferramenta e utilizados nas próximas práticas (os demais arquivos podem ser descartados).

2.2 USO DA FERRAMENTA

O diretório “/Ferramentas/Extrator de Características” disponibilizado no [link](#) contém a ferramenta em Java para realizar a extração das características. Há também um arquivo “readme” de como utilizar a ferramenta e um arquivo em formato .pdf contendo as propriedades de cada extrator, incluindo a categoria e o número de características a serem extraídos.

3 PRÁTICA 2

Nesta prática será utilizada a plataforma Google Colab para realizar os experimentos envolvendo aprendizado supervisionado, por meio de classificadores tradicionais. Ao final, pretende-se obter as melhores combinações de extratores e classificadores para cada conjunto de dados selecionado. Exemplos de códigos para execução dos experimentos encontram-se disponíveis no [link](#).

3.1 ENUNCIADO

A partir dos arquivos de características (em formato .arff) obtidos na Prática 1, realize experimentos com os classificadores. Além disso, para cada um deles considere diferentes cenários: i.e. alterando os parâmetros de entrada para os classificadores (quando possível), bem como as técnicas de normalização. Analise o desempenho de cada classificador em cada conjunto de dados e cenário considerado. Para tanto, obtenha as respectivas acurácias, gráficos e métricas de avaliação.

Observação:

- Armazene todos os resultados obtidos pelos respectivos extratores e classificadores (em diferentes cenários) para que os mesmos possam ser analisados, selecionados e apresentados posteriormente no relatório final.

4 PRÁTICA 3

Nesta prática será utilizada a plataforma Google Colab para realizar os experimentos envolvendo aprendizado não-supervisionado, por meio do método de agrupamento k -means. Ao final, pretende-se obter o agrupamento das amostras do conjunto de dados selecionados. Exemplos de códigos para execução dos experimentos encontram-se disponíveis no [link](#).

4.1 ENUNCIADO - SUGESTÕES

A partir dos arquivos de características (em formato .arff) obtidos na Prática 1, realize experimentos com o algoritmo de agrupamento k -means. Além disso, para cada um dos extratores, considere diferentes cenários: i.e. alterando o valor de k , a técnica de normalização, bem como a utilização da técnica PCA. Analise o desempenho de cada extrator em cada conjunto de dados e cenário considerado, observando a disposição no espaço de características e entendendo as técnicas implementadas.

Observação:

- Armazene todos os resultados obtidos pelos agrupamentos realizados (em diferentes cenários), para que os mesmos possam ser analisados, selecionados e apresentados posteriormente no relatório final.

5 PRÁTICA 4

Nesta prática será utilizada a plataforma Google Colab para realizar os experimentos envolvendo aprendizado semi-supervisionado, por meio de classificadores tradicionais combinados com o método Label Propagation. Ao final, pretende-se obter a familiarização com essa forma de aprendizado. Exemplos de códigos para execução dos experimentos encontram-se disponíveis no [link](#).

5.1 ENUNCIADO

A partir dos arquivos de características (em formato .arff) obtidos na Prática 1, realize experimentos com alguns classificadores. Além disso, para cada um deles considere diferentes cenários: i.e. alterando os parâmetros do método Label Propagation e as técnicas de normalização. Analise o desempenho de cada classificador semi-supervisionado em cada conjunto de dados e cenário considerado. Para tanto, obtenha as respectivas acurácias, gráficos e métricas de avaliação. Realize comparações entre as abordagens de aprendizado semi-supervisionada e supervisionada (obtidas na Prática 2).

Observação:

- Armazene todos os resultados obtidos pelos respectivos extratores e classificadores (em diferentes cenários) para que os mesmos possam ser analisados, selecionados e apresentados posteriormente no relatório final.

6 PRÁTICA 5

Nesta prática será utilizada a plataforma Google Colab para realizar os experimentos envolvendo aprendizado ativo, por meio de classificadores tradicionais. Ao final, pretende-se obter o entendimento básico dessa forma de aprendizado e a proposta de novos métodos de seleção de amostras informativas. Exemplos de códigos para execução dos experimentos encontram-se disponíveis no [link](#).

6.1 ENUNCIADO

A partir dos arquivos de características (em formato .arff) obtidos na Prática 1, realize experimentos com alguns classificadores. Nesse experimento utilize as configurações referentes aos melhores cenários obtidos nas práticas anteriores (e.g. técnica de normalização, parâmetros, entre outros). Observe a evolução do processo de aprendizado (acurácias) dos classificadores ao longo das iterações de aprendizado e das interações do especialista. Realize comparações entre abordagens de aprendizado ativo e supervisionadas tradicionais (e.g. obtidas na Prática 2). Para tanto, verifique em que iteração do processo de aprendizado ativo é possível obter valores de acurácias similares aos obtidos por meio do processo de aprendizado supervisionado tradicional. Compare as quantidades de amostras anotadas necessárias para cada tipo de abordagem de aprendizado atingir os mesmos valores de acurácia. Ao final, apresente a proposta de uma estratégia para seleção de amostras mais informativas.

Observação:

- Armazene todos os resultados obtidos pelos respectivos extratores e classificadores (considerando as configurações referentes aos melhores cenários), para que os mesmos possam ser analisados, selecionados e apresentados posteriormente no relatório final.