

PMR 3508 – Aprendizado de Máquina e Reconhecimento de Padrões

Prof. Fabio G. Cozman

Objetivo: Apresentar técnicas básicas de aprendizado de máquina e indicar suas aplicações, em particular em reconhecimento de padrões e processamento de imagens.

Avaliação: Duas provas e trabalhos: $NF = (P1+2*P2+T)/4$.

T = média de atividades práticas de aprendizado de máquina.

Material:

Livro texto: G. James, D. Witten, T. Hastie, R. Tibshirani. *An Introduction to Statistical Learning*, Springer 2013. [Disponível em <http://www-bcf.usc.edu/~gareth/ISL/>.]

Livros adicionais: D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge 2012;

K. R. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012;

C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006.

Desenvolvimento: Python+Panda (Data Camp: Intro to Python for Data Science); Jupyter, scikit-learn, Keras+Tensorflow.

Dados: UCI repository (<https://archive.ics.uci.edu/ml/>); Kaggle (www.kaggle.com/).

Programa:

Introdução: a tecnologia de aprendizado de dados e de grandes bases de dados.

Definições básicas (tipos de aprendizado, tratamento de dados, avaliação); exemplo: *k-nearest-neighbor* (kNN).

Probabilidade (variáveis discretas); exemplo: *Naïve Bayes* (aplicação em detecção de spam).

Avaliação cruzada e medidas de desempenho (acurácia, ROC, AUC).

Probabilidade (variáveis contínuas).

Estimação MLE: modelos hierárquicos; exemplo: redes Bayesianas (aplicação na base *adult*).

Dados faltantes e o algoritmo EM.

Estimação Bayesiana; teoria da decisão Bayesiana.

Métodos de estimação MCMC.

Estatística frequentista: estimação e testes.

Testes para comparação de métodos; técnicas de *bootstrap*.

Regressão linear; MSE (vício+variância); análise de erro de classificação.

Regressão não-linear e não-paramétrica.

Regressão *ridge* e *lasso*; problemas com muitos atributos.

Regressão logística e discriminante linear/quadrático.

SVMs e *kernels*.

Árvores de classificação: algoritmos C4.5, C5.0, CART.

Bagging e *random forests*.

Boosting: algoritmo AdaBoost.

Métodos de agrupamento: *K-means*, dendrogramas.

PCA (NMF); aplicação em sistemas de recomendação.

Modelos latentes (LDA).

MDPs e aprendizado por reforço

Redes neurais e *deep learning*.

Deep reinforcement learning.