



CC-226

Introdução à Análise de Padrões

Prof. Carlos Henrique Q. Forster

Plano de Curso
E Primeira Tarefa



Professor

- Prof. Carlos Henrique Q. Forster
- Instituto Tecnológico de Aeronáutica
- Sala 121 IEC
- ramal 5981
- e-mail:
- Página do curso:
<http://www.comp.ita.br/~forster/CC-226>



Plano de aulas – Tópicos

- Introdução – Validação da Classificação
- Distribuições de Probabilidade e Modelos Lineares
- Inferência Bayesiana e Teoria da Decisão
- Análise de Componentes
- Estimação não-paramétrica
- Clustering
- Estimação de Parâmetros
- Redes Bayesianas
- Outros modelos de aprendizagem de máquina



Avaliação

- 4 trabalhos
- Nota do primeiro bimestre:
 - Trabalho 1 – Análises utilizando Weka
 - Peer-review do trabalho 1
 - Trabalho 2 – Análises com implementações em linguagem de programação
 - Peer-review do trabalho 2



- Nota do segundo bimestre:
 - Trabalho 3 – Análises com métodos avançados e implementação do trabalho final
 - Peer-review do trabalho 3

- Nota do exame:
 - Trabalho final – Escrita de artigo padrão de conferência internacional (pode ser em Português), continuação do trabalho 3.
 - Seminário – apresentação do trabalho final



Ferramentas

- ❑ Weka
- ❑ Matlab + statistics toolbox ou Octave
- ❑ Linguagens C, Java, Python ou R
- ❑ Ferramentas do Python: numpy/scipy, IPython, Pandas, statsmodels, matplotlib, sympy, scikit-learn, numba, blaze, bokeh, PyMC
- ❑ OpenCV, NLTK, NetworkX
- ❑ Mondrian ou outros pacotes de visualização



Livros

- ❑ DUDA, HART e STORK - Pattern Classification
- ❑ DEVORE - Probabilidade e Estatística para Engenharia e Ciências
- ❑ TEODORIDIS e KOUTROMBAS – Pattern Recognition
- ❑ ALPAYDIN - Introduction to Machine Learning
- ❑ MITCHEL - Machine Learning
- ❑ BISHOP - Pattern Recognition and Machine Learning
- ❑ HAN e KAMBER - Data Mining Concepts and Techniques
- ❑ WITTEN e FRANK – Data Mining



Conferências

- ❑ Computer Vision and Pattern Recognition
- ❑ International Conference on Machine Learning
- ❑ Neural Information Processing Systems
- ❑ Intl. Conference on Knowledge Discovery and Data Mining
- ❑ ACM SIGKDD
- ❑ SIAM Data Mining Conference



Revistas

- ❑ Pattern Analysis and Applications (Springer)
- ❑ IEEE Transactions on PAMI
- ❑ Pattern Recognition (Elsevier)
- ❑ IEEE Transactions on Neural Networks
- ❑ IEEE TKDE
- ❑ Journal of Machine Learning Research www.jmlr.org
- ❑ Journal of the American Statistical Association
- ❑ ACM Transactions on KDD
- ❑ Data Mining and Knowledge Discovery (Elsevier?)



Fontes de Dados

- UCI Repository:
<http://www.ics.uci.edu/mlearn/MLRepository.html>
- UCI KDD Archive:
<http://kdd.ics.uci.edu/summary.data.application.html>
- Statlib: <http://lib.stat.cmu.edu/>
- <http://www.kdnuggets.com/datasets/index.html>

- Exemplos do Weka



Proposta de Calendário

Sem.	Data	Prazo	Conteúdo
1	29/7		Aula - introdução e validação da classificação
2	5/8		Reservado para trabalho1
3	12/8	Trabalho1	Aula - distribuição probabilidades e regressão
4	19/8	Peer-review1	Aula - Inferência bayesiana e teoria da decisão
5	26/8		Aula - Análise de componentes
6	2/9		Reservado para trabalho2
7	9/9	Trabalho2	Aula - Clustering
8	16/9	Peer-review2	Aula - Estimção não-paramétrica

Sem.	Data	Prazo	Conteúdo
9	30/9		Reservado para trabalho3
10	7/10		Reservado para trabalho3
11	14/10	Trabalho3	Aula - Estimção de parâmetros
12	21/10	Peer-review3	Aula - Redes Bayesianas
13	28/10		Aula - Outros modelos de aprendizagem
14	4/11		Reservado para preparação dos seminários
15	11/11	Seminários	Seminários do trabalho final
16	18/11		Seminários do trabalho final



Trabalho1

- Para 12/Agosto
- Construção de classificadores
 - Grupos de no máximo 3 pessoas
 - Aprender a utilizar e explorar o pacote Weka
 - <http://www.cs.waikato.ac.nz/ml/weka/>
 - Familiarizar-se com os datasets da UCI
 - <https://archive.ics.uci.edu/ml/datasets.html>



Trabalho1

- Construir e avaliar 5 classificadores para 5 problemas definidos a partir de conjuntos de dados. 2 deles dos exemplos do Weka e 3 escolhidos no repositório da UCI. Utilizar um tipo diferente de classificador para cada problema.
- Emitir relatório explicando/definindo cada um dos 5 problemas/datasets, qual classificador foi criado e como obter a estrutura do classificador a partir do Weka, avaliar os resultados da classificação explicando 5 formas de avaliação ou tipos de índices diferentes fornecidos pelo Weka. Traçar e analisar a curva ROC em cada caso. Desenhar a estrutura do classificador quando possível.