

CC-226 Introdução à Análise de Padrões

Apresentação do Curso

Carlos Henrique Q. Forster¹

¹Divisão de Ciência da Computação
Instituto Tecnológico de Aeronáutica

25 de fevereiro de 2008



1

O Curso

- Professor - Contato
- Plano de Aulas
- Ferramentas
- Livros adotados
- Pré-Requisitos
- Escopo
- Que tipo de problema vamos resolver?
- Calendário
- Avaliação
- Trabalho do 2o bimestre - para 18/junho
- Onde pesquisar



Prof. Carlos Henrique Q. Forster
Instituto Tecnológico de Aeronáutica

- Sala 121 IEC
- ramal 5981
- e-mail: forster@ita.br



Dedicação ao curso

3-0-6

- Horas em sala de aula: 3
- Aulas de laboratório: 0
- Horas de estudo em casa: 6



- MATLAB ou OCTAVE
- Linguagem C ou Java ou Python
- Weka ou software de análise estatística
- Google ;)



- DUDA, HART e STORK - Pattern Classification
- DEVORE - Probabilidade e Estatística para Engenharia e Ciências
- MITCHEL - Machine Learning
- BISHOP - Pattern Recognition and Machine Learning
- HAN e KAMBER - Data Mining Concepts and Techniques

Material adicional: LINKS, PAPERS e Capítulos.



- Cálculo em múltiplas variáveis
- Álgebra linear
 - Mínimos quadrados
 - Autovalores e autovetores
- Métodos numéricos e de otimização
- Probabilidades e Estatística
- Algoritmos e Estruturas de dados

É assumida familiaridade com assuntos do curso de Inteligência Artificial: Métodos de busca e otimização, representação do conhecimento, redes neurais, incerteza e lógica dedutiva e indutiva.



- Cálculo em múltiplas variáveis
- Álgebra linear
 - Mínimos quadrados
 - Autovalores e autovetores
- Métodos numéricos e de otimização
- Probabilidades e Estatística
- Algoritmos e Estruturas de dados

É assumida familiaridade com assuntos do curso de Inteligência Artificial: Métodos de busca e otimização, representação do conhecimento, redes neurais, incerteza e lógica dedutiva e indutiva.



- Curso introdutório - nos preocuparemos com abordagens mais clássicas do que tópicos avançados.
- Mais conceitual.
- É necessário entendimento de como e quando as técnicas apresentadas funcionam e, quando não são adequadas, o porquê e como se contorna o problema.
- Os algoritmos costumam ser fáceis de implementar. É mais importante preparar os dados de entrada numa forma adequada ao problema que se quer resolver, validar os resultados e comparar com outras técnicas.



Que tipo de problema vamos resolver?

- A partir de um conjunto de dados, obter uma equação que generalize seu modelo e permita prever valores não observados.
- A partir de dados e de medidas sobre um objeto em particular, discriminar esse objeto como sendo de uma classe específica correspondente aos dados.
- Detectar instâncias cujos dados não se comportam como a maioria das instâncias.
- Comparar conjuntos de dados.
- Estimar valores de variáveis não-observadas, dado um modelo e um conjunto de dados.
- Encontrar um conjunto mínimo de atributos e que melhor caracterizam um conjunto de dados.
- Avaliar o risco e a validade de uma classificação.
- Identificar grupos com instâncias de comportamento similar em um conjunto de dados.
- Identificar distribuições e/ou regras que governam o comportamento dos dados.



Calendário (aproximado, estimado, possível)

	Data	Tópico		Data	Tópico
1	27/2	Descrição	9	30/4	Superfícies de separação
2	5/3	Revisão probabilidades	10	7/5	Clustering e EM
3	12/3	Regressão paramétrica	11	14/5	Análise de componentes
4	19/3	Regressão não-param.	12	21/5	Testes estatísticos
5	26/3	Inferência bayesiana	13	28/5	Avaliação de classific.
6	2/4	Teoria da decisão	14	4/6	Árvores e visualização
7	9/4	Estimação de parâmetros	15	11/6	??
8	16/4	**Prova**	16	18/6	**Seminário**

Feriados: 21/março (sexta) 21/abril (segunda) 1/maio (quinta) 22/maio (quinta)



Avaliação - 1o bimestre

- Lista de exercícios (2 pontos) (individual)
- Implementações (4 pontos) (individual)
- Prova (4 pontos)

Avaliação - 2o bimestre

- Trabalho final (implementação, relatório, apresentação) (10 pontos) (grupos)



Trabalho do 2o bimestre - para 18/junho

- Escolher um dos papers apontados pelo professor ou outro paper escolhido por vocês e aprovado pelo professor. (até 16/abril) Terá preferência ao tema o grupo que escolher primeiro (por e-mail por exemplo).
- Montar a equipe de até 3 pessoas. (até 16/abril)
- Entregar em um CD o conjunto de dados para teste, o código-fonte em MATLAB, C, Java ou Python, o código executável, o relatório e a apresentação. (até 18/junho)
- Agendar sua apresentação (todos os membros do grupo devem apresentar) para 18/junho ou 25/junho.



Natureza do trabalho

- Basear-se no método ou aplicação encontrado em um paper.
- Selecionar uma base de dados para análise ou criar sua própria base.
- Estabelecer questões a se analisar nesses dados.
- Formular de forma específica ao problema e implementar técnicas apresentadas no curso.
- Aplicar as técnicas, obter resultados, validar esses resultados e fornecer uma avaliação do quanto esses resultados são seguros, confiáveis e significativos.
- Comparar as técnicas utilizadas e os relatos da literatura e identificar deficiências e oportunidades de melhorias nos métodos utilizados.



● Conferências

- Compute 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

● Jornais

- IEEE PAMI
- Pattern Recognition
- Neural Networks
- IEEE TKDE
- Journal of the American Statistical Association
- ACM Transactions on KDD
- Data Mining and Knowledge Discovery

