



Data Mining

Mestrado SIAD

2007 – 2008

Programa

Introdução ao Data Mining	Tipos de dados. Tarefas e problemas de Data Mining. Ferramentas. Recolha e organização dos dados. Casos omissos e seu tratamento. "Outliers", sua identificação e tratamento. Transformação de variáveis. Exemplos. Revisão: a hipótese nula e os testes de hipóteses.
Regressões Lineares e não Lineares	Validade. Transformações. Escolha de variáveis, análise de resíduos, variabilidade explicada. Casos de não linearidade. Algoritmos para regressões não lineares. Aplicações empresariais dos métodos de regressão: marketing.
Regressões Logísticas	Da regressão linear à regressão logística. "odds" vs probabilidades. Regra de Bayes e ganho em informação. Forma sistémica das regressões. Algoritmos discriminantes. Fronteiras, construção de índices predictivos. Custos associados a cada tipo de erro. Aplicações empresariais: banca.
Arquitectura das redes neuronais	Forma sistémica, arquitectura, algoritmos para aprendizagem, tipos de função a minimizar, tipos de função de transferência. Discussão.
Prática de modelação	Construção de modelos de índices predictivos de falência com algoritmos lineares de Fisher e com regressões logísticas. Interpretação e tratamento posterior. <i>Credit Scoring</i> . Gestão de Risco. Aplicações em marketing.
Retro-Propagação	A regra de Delta e a regra de Delta Generalizada. Treino dos MLP, mínimos locais, alternativas.
Prática de modelação	Construção de modelos para previsão de falência com Perceptrões de camadas múltiplas. Aferição e medição da efectividade, tipos de erro, "hold out samples".
Análise de grupos	Grupos de casos e de variáveis. Mapeamento de casos, de variáveis e de estruturas. A vizinhança, suas várias definições e aplicação a cada caso. Os algoritmos hierárquicos e sua caracterização. Exemplos: marketing.
Redes de Kohonen	Mapas topológicos, redução dimensional, regra de Hebb, auto-organização. Aplicações e diferenças em relação a outros algoritmos. Identificação de grupos de empresas em "distress" financeiro. Risco de Crédito. Detecção de fraudes.
Probabilidades, Informação e Incerteza	Incerteza. Algoritmos para a estimação da informação incompleta e sua aplicação. Exemplos. Regra de Bayes. Mecanismos aleatórios e tipos de dados. Aplicações: análise financeira, projectos de investimento.
Árvores de Decisão	Modelos com estrutura hierárquica, limitações e aplicação correcta deste tipo de algoritmos, diferentes tipos de critérios, algoritmos para atributos contínuos, diferentes tipos de "pruning". Visão crítica. Uso do algoritmo tipo C5 para redução da complexidade e avaliação do grau de controlo do gestor em projectos de investimento.
Análise de Afinidades	Regras associativas. Limitações e diferentes tipos. Algoritmos a priori. Dados categoriais, lógicos ou contínuos. Aplicações a cabazes de produtos. Intervalos de confiança.
Revisões práticas de modelação	Trabalho individual acompanhado (exercícios de construção de índices predictivos, de redução de complexidade, estudos de mercado, definição de cabazes).