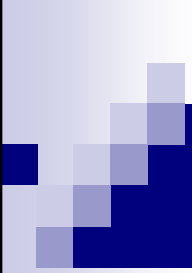


Utilização do SAS Enterprise Miner

V 3.0, V.Lobo, EN/ISEGI, 2005



Datamining em SAS

Victor Lobo

Problema 1: Iris

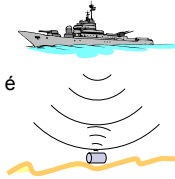
- A flor Iris tem várias variantes, 3 das quais são:
 - 1 -Iris Setosa
 - 2 -Iris Versicolour
 - 3 -Iris Virginica
- Para 50 flores de cada uma das variantes foram medidas 4 características (medidas em cm)
 - Largura da pétala
 - Comprimento da pétala
 - Largura da Sépala
 - Comprimento da Sépala
- **Questão:**
 - É possível determinar a variante a partir desses 4 parâmetros ?



Iris Setosa

Problema 2: Encontrar minas

- Um navio pode procurar minas pousadas no fundo do mar com o seu sonar
 - O sonar envia um pulso de som que é reflectido pela mina. Esse eco é decomposto num espectro com 60 bandas de frequência.
 - O eco produzido por uma rocha é semelhante, mas ligeiramente diferente
- **Questão:**
 - É possível distinguir minas de rochas a partir do eco ?



Problema 3: Artificial...

- Seja $f(x,y,z)=2.1 x + y x z -1$
 - Obtenha 100 conjuntos de pontos (x,y,z) gerados aleatoriamente no intervalo $[0,10]$
 - Calcule o valor de f para cada um desses pontos
 - Treine uma rede neuronal para calcular a função
 - Calcule o erro entre a rede e a função verdadeira. Teste com novos pontos

SAS Enterprise Miner

- Módulo de SAS para datamining
 - Menu Solutions → Analysis → Enterprise Miner
- Inclui métodos de previsão e clustering
 - Redes Neuronais
 - Perceptrão Simples, MLP, RBF, SOM
 - Árvores de decisão
 - Classificação por protótipos
 - Clustering Hierárquico
 - Regras de associação
 - Regressão (clássica)

Enterprise Miner

- Inclui ferramentas de manipulação de dados
 - Partição em treino/validação/teste
 - Amostragem
 - Remoção de outliers
 - Componentes principais
- Inclui ferramentas para comparação de resultados, geração de "reports", código Sas
 - Score
 - Assessment
 - Reporter
 - SAS code generator

Utilização do SAS Enterprise Miner

V 3.0, V.Lobo, EN/ISEGI, 2005

Importação de dados

- Os dados têm que estar numa biblioteca SAS
- Importar dados
 - Fora do Enterprise Miner fazer:
 - File → Import data (e seguir o wizard)
 - Formatos suportados
 - Excel (XLS)
 - Texto simples (TXT, com tabs ou CVS, com vírgulas)
 - dBase (DBF)

Cada projecto é um “DIAGRAM”

- Criar usando **New → Diagram**
- Definir origem dos dados (já têm que estar em SAS), com **Input Data Source**
 - Definir tipo de cada atributo (measurement)
 - Definir função de cada atributo (model role)
 - Explorar os diversos tags...
- Particionar o conjunto de dados com **Data Partition**.
 - Explorar os diversos tags...

Inserir o objecto **Neural Network**

- No tag General, definir se queremos classificação ou regressão definindo **Model selection criteria**.
 - Usar MLP em modo “basic”
- Correr o modelo com **Run**.
 - Ver os resultados: classification error, pesos, classificações feitas