



Exercício sobre Machine Learning

ROTEIRO:

1. Ler o material disponibilizado aos alunos: (livro disponível na Biblioteca)
Cap. 4 do livro Sistemas Inteligentes – Solange Rezende (ed)

Seria interessante que o aluno também consultasse o material abordado na aula sobre aprendizado de máquina que está disponível na internet, para a realização deste trabalho.

Introdução: Conceitos de Inteligência, Aprendizado, Sistema Especialista, Aquisição de Conhecimentos e Aprendizado de Máquina (Machine Learning). –Transparências
<http://www.inf.unisinos.br/~osorio/sistadap.html> (Aula01)

2. Leia o exemplo (discussão inicial), siga as instruções dadas abaixo, e responda a questão proposta de acordo com o solicitado.

Considere o seguinte exemplo:

Coleção de Dados-Exemplo

Estes dados serão usados apenas para uma demonstração inicial sobre o problema.

Atributo1 = Frequência dos Batimentos Cardíacos

Atributo2 = Respiração [não, sim]

Classe = 0: Morto

1: Vivo_Descansado

2: Vivo_Cansado (sob exercício)

Dados Históricos (registro de ocorrências passadas)

Atributo1 (freq. card)	Atributo2 (respiração)	Classe
80	sim	1
90	sim	1
0	não	0
130	sim	2
40	sim	1

Regras

Supondo que um tenha sido contratado como “estagiário” e deva proceder a um exame de uma pessoa, realizando a classificação desta pessoa, nas 3 classes descritas acima, onde a única informação disponível é um registro histórico de classificações prévias feitas pelo seu chefe... como ele deveria proceder?

Regras criadas “a mão” pelo estagiário para “automatizar” o processo:

```
IF Respiração = não  
  THEN Classe = Morto
```

```
IF Respiração = sim AND Batimentos < 100  
  THEN Classe = Vivo_Descansado
```

```
IF Respiração = sim AND Batimentos >= 100  
  THEN Classe = Vivo_Cansado
```

Para pensar...

Após ter criado suas regras, aparece a necessidade de fazer uma nova classificação...

```
Atrib1  Atrib2  Classe  
50      não     ?
```

Como será que o sistema proposto acima reagiria? Como será que o estagiário poderia melhorar o sistema dele de modo a minimizar as classificações incorretas de novos casos?

Existe aqui uma preocupação para com a “GENERALIZAÇÃO”, ou seja, identificar as regras mais genéricas possíveis de modo a evitar possíveis erros. Obviamente que isto ficaria mais fácil se tivéssemos mais dados disponíveis sobre o problema para poder criar regras mais adequadas.

Sugestão...

- Para tabelas maiores: use o Excel, por exemplo
- Tente achar uma "ordenação" nos dados
- Dados com 2 ou 3 atributos de entrada (2D ou 3D):
 Tente plotar os dados: use o gnuplot, por exemplo

DEFINIÇÃO DA TAREFA PRÁTICA:

Coleção de Dados e Regras

Obter uma representação dos dados de aprendizado na seguinte forma:

```
IF atributo operador valor [And/Or ...] THEN Classe
```

```
Atributo1 = Temperatura  
Atributo2 = Pressão  
Classe = 0: OK  
        1: Problema
```

Dados de Aprendizado (para a indução de regras)

```
Atrib1  Atrib2  Classe  
100     50     1  
80      60     1  
85      40     1  
90      60     1  
20      20     0  
90      40     1
```

30	25	0
20	25	0
100	60	1
40	25	0
90	50	1
85	50	1
30	20	0
80	40	1
100	45	1
30	15	0
85	45	1
80	50	1
20	15	0
40	15	0
90	45	1
85	60	1
40	20	0
100	40	1
80	45	1

TAREFAS:

- 1) Obter duas representações distintas destes dados representados por regras da seguinte forma: (duas possíveis soluções para a automação desta tarefa através do uso de regras)

IF atributo operador valor [And/Or ...] THEN Classe

=> Para cada um dos conjuntos de regras:

- 2) Obter a matriz de confusão, USANDO AS REGRAS que você criou como classificador dos seguintes dados (dados de teste) abaixo:

Dados de Teste

Atrib1	Atrib2	Classe
56	39	0
85	60	1
20	25	0
90	45	1
34	18	0
81	23	0
26	5	0
62	38	0
100	40	1
60	37	0
30	15	0
50	26	1
0	0	1
30	20	0
92	19	0

- 3) Calcular as seguintes medidas:
 - 3.1. Erro (taxa de classificação incorreta)
 - 3.2. Precisão do seu classificador (acurácia)
 - 3.3. Distribuição de classes em ambos conjuntos de dados (dados originais e dados de teste)
 - 3.4. Erro majoritário para o conjunto de dados de teste
- 4) Indicar qual dos classificadores (entre os 2 conjuntos de regras que você criou) que você considera mais adequado para ser adotado. Justifique sua opção e descreva qual o critério que você usou para realizar esta opção.

- 5) Considerando os dados abaixo (aprendizado e teste), indicar um conjunto de regras de classificação para os mesmos e a respectiva matriz de confusão obtida com a aplicação de seu classificador. Problema: 3 entradas e uma saída binária (0 ou 1).

Aprendizado:

```
0.3 0.1 0.4 0
0.1 0.1 0.8 0
0.2 1.0 0.0 0
0.1 0.9 0.7 1
1.0 0.0 0.3 0
0.6 0.3 1.0 1
1.0 1.0 0.4 1
1.0 1.0 0.6 0
```

Teste:

```
0.0 0.4 0.1 0
0.2 0.2 1.0 0
0.4 0.6 0.2 0
0.2 0.8 0.6 1
0.9 0.3 0.2 0
0.8 0.1 0.9 1
0.9 1.0 0.0 1
0.7 0.7 0.9 0
```

Bom trabalho!

Atenção:

- ⇒ **Entregar para o professor o trabalho (por e-mail ou pessoalmente) até a próxima aula.**
- ⇒ **Os dados deste arquivo estão disponíveis em <http://inf.unisinos.br/~osorio/iajogos/ml>**