

Manual do Usuário

Pablo Dalbem de Castro RA 038499
George Barreto Pereira Bezerra RA 003030

Tema: Aprendizado em Redes Bayesianas

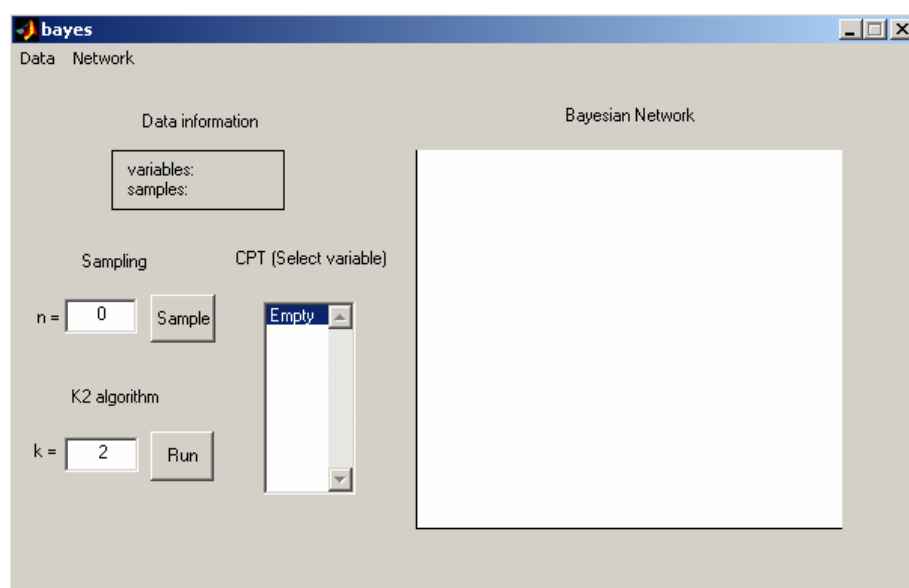
O software BayesNet é implementado em linguagem Matlab e possui uma interface gráfica de fácil utilização. BayesNet utiliza o algoritmo K2 para inferir redes bayesianas a partir de um conjunto de dados discretos inseridos pelo usuário. O software inclui também uma funcionalidade adicional que permite gerar amostras a partir de uma rede já construída. Este manual explica como utilizar o programa para que você crie a sua própria rede bayesiana.

1. Instalando o software

Para instalar o software basta copiar o programa para o diretório desejado e ele já estará pronto para ser utilizado. Dentro do diretório principal, o diretório *projeto*, estão localizados todos os arquivos necessários para que o programa funcione. Há dentro de *projeto* mais dois diretórios: *data* e *network*. O diretório *data* é utilizado para armazenar dados do usuário a serem utilizados na inferência de uma rede ou então dados gerados por uma amostragem de uma rede já construída. O diretório *network* serve para armazenar redes já inferidas ou redes quaisquer que vão ser utilizadas para amostragem.

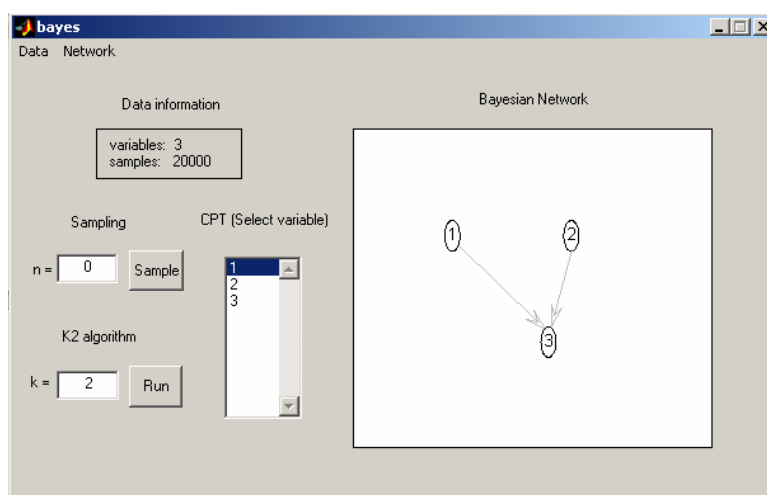
2. Criando sua primeira rede

Essa seção ensina como criar a sua primeira rede bayesiana. Para rodar o programa, vá ao diretório *projeto* dentro do Matlab e digite *bayes*. A seguinte janela aparecerá na tela:



Agora você deve carregar os dados para gerar a sua rede. Vá ao menu *Data* e selecione *Load*. Esta opção explorará o diretório *data*, o qual possui uma série de arquivos como exemplos. Selecione o arquivo *heckerman*. As informações do arquivo devem aparecer no painel *Data information*, onde *variables* significa o número de variáveis e *samples* o número de amostras contidas nos dados.

Para rodar o algoritmo, vá ao painel *K2 algorithm*. Lá você encontrará a opção de configurar o parâmetro *k*, que significa o número máximo de pais que uma variável pode ter na rede. Quanto maior este valor, maior a complexidade que uma rede pode assumir. Deixe este parâmetro com seu valor padrão por enquanto. Clique agora no botão *Run* para que o algoritmo comece a construir a rede. O painel *Bayesian Network* deve mostrar a evolução da topologia da rede durante a execução do algoritmo. Por fim, a janela do programa deve aparecer como na figura abaixo, onde a topologia final da rede é mostrada no painel *Bayesian Network*:



Para visualizar as tabelas de probabilidade, basta selecionar a variável desejada no painel *CPT*. Clicando na variável 3, por exemplo, a seguinte janela deve aparecer na tela:

CPT of variable 3			
		1	2
1	1	0.2057	0.7957
1	2	0.5787	0.4227
2	1	0.3617	0.6337
2	2	0.6377	0.3637

Esta janela apresenta os valores de probabilidade para todas as combinações de estado pais-filho. Note que as variáveis são binárias. Os valores binários 11, 12, 21 e 22 representam o estado dos pais, sendo a variável 1 à esquerda e a variável 2 à direita (variáveis de menor índice sempre ficam à esquerda das variáveis de maior índice). Na primeira linha da tabela, são mostrados os estados da variável filha (variável 3 da rede). Por exemplo, a probabilidade que a variável 3 assumo o valor 1 dado que as variáveis 1 e 2 assumem

ambas valores 1 é 0,205. A probabilidade que a variável 3 assuma valor 2 dado que seus pais (variáveis 1 e 2) assumem valor 2 e 1, respectivamente, é 0,639.

Para salvar a rede obtida basta ir ao menu *Network*, selecionar a opção *save* e digitar o nome desejado para o arquivo.

3. Gerando amostras a partir de uma rede

Para gerar amostras a partir de uma rede basta ir ao menu *Network* e selecionar uma rede qualquer através do menu *load*. (É possível selecionar também a rede que você acabou de salvar durante a seção anterior.)

Selecione o arquivo *cloudy.mat*. A estrutura da rede deve aparecer na tela. Vá agora ao painel *sampling* e no campo *n* digite o número de amostras que você deseja gerar, por exemplo, 1000. Para gerar as amostras basta clicar em *sample*. Os dados gerados serão mostrados na tela de comandos do Matlab. Para salvar os dados vá ao menu *data* e selecione *save*. Esses dados podem agora ser utilizados para gerar uma rede bayesiana através do algoritmo K2.

4. Dados

Para gerar a sua rede bayesiana, é necessário que o usuário entre com os dados. Estes devem ser discretos (não necessariamente binários) e é desejável que possuam um número elevado de amostras.

Os dados de entrada são matrizes, onde cada coluna representa uma variável e cada linha, uma amostra. Um exemplo de conjunto de dados é mostrado abaixo:

2	1	5	4
1	1	5	4
1	2	2	1
2	2	1	3
1	2	3	5
1	1	4	1

Os dados na tabela possuem 4 variáveis e 6 amostras. Repare que as duas primeiras variáveis são binárias, isto é, assumem apenas dois valores, 1 e 2. Uma observação importante aqui é que o valor zero não é permitido. Os valores que uma variável pode assumir devem ser sempre inteiros e começando de 1. As duas últimas variáveis também são discretas, mas variam de 1 a 5, isto é, possuem 5 estados.

5. Acessando a rede bayesiana

Ter acesso direto às redes bayesianas sem a interface do programa é simples. Vá, através da tela de comando do Matlab, ao diretório *network*, onde se encontra a rede desejada. Carregue a rede através do comando *load*. Assim, três variáveis devem ser carregadas no *workspace* do Matlab: *S*, *tables* e *typ*. A variável *S* contém a estrutura da rede. É uma matriz quadrada em que o valor 1 representa um arco e o valor 0 a ausência de conexões. A tabela

abaixo dá um exemplo de uma estrutura de rede. Para cada arco da matriz, a linha representa a variável filha e a coluna, a variável pai. Neste exemplo, a variável 1 é pai das variáveis 2 e 3.

0	0	0
1	0	0
1	0	0

A variável *tables* contém as tabelas de probabilidade. Para acessar uma tabela de probabilidade, basta digitar *tables.cpt(i)*, onde *i* é o índice da variável desejada. Os valores de probabilidade devem aparecer na tela.

A variável *typ* corresponde a um vetor com o “tipo” de cada variável, isto é, com o número de valores discretos que cada variável pode assumir. Por exemplo, um vetor *typ* = [4 2 2] representa uma primeira variável com 4 valores discretos e as outras duas binárias.

Para construir manualmente a sua própria rede bayesiana, o usuário deve definir as variáveis *S*, *tables* e *typ* e salva-las em um arquivo para que possam ser carregadas pelo programa.