Universidade de São Paulo – USP
Instituto de Ciências Matemáticas e de Computação – ICMC
Departamento de Ciências de Computação – SCC

SCC-5949 – Inteligência Artificial II

Professor Gustavo Batista
gbatista@icmc.usp.br

# Projeto – Redes Bayesianas

**Data de entrega: 22/06**

**Este projeto é individual. A entrega deve ser realizada via TIDIA, fazendo *upload* na pasta "projeto" na ferramenta escaninho.**

O objetivo deste projeto é implementar uma rede bayesiana calculando as probabilidades marginais e condicionais a partir de uma versão do conjunto de dados de *benchmark* *Contraceptive Method Choice* (CMC).

O domínio de aplicação deste projeto é a escolha de métodos contraceptivos. O conjunto de dados possui 1473 registros e nove atributos listados a seguir:

1. Wife's age (categorical) **1=** [0,26), **2 =** [26,32), **3 =** [32,39), **4 =** [39,+∞)
2. Wife's education (categorical) **1**=low, **2**, **3**, **4**=high
3. Husband's education (categorical) **1**=low, **2**, **3**, **4**=high
4. Number of children ever born (categorical) **0**=0, **1**=1, **2**=2, **3**=4, **4**=[4,5], **5**=[6,+∞)
5. Wife's religion (binary) **0**=Non-Islam, **1**=Islam
6. Wife's now working? (binary) **0**=Yes, **1**=No
7. Husband's occupation (categorical) **1**, **2**, **3**, **4**
8. Standard-of-living index (categorical) **1**=low, **2**, **3**, **4**=high
9. Media exposure (binary) **0**=Good, **1**=Not good
10. Contraceptive method used (class attribute) **1**=No-use, **2**=Long-term, **3**=Short-term

Os atributos de um a nove são atributos preditivos e o atributo dez é classe. Nesse conjunto de dados, o objetivo final é predizer o atributo classe por meio dos atributos preditivos.

## Parte I – Proposta de uma rede bayesiana e cálculo das probabilidades a partir de dados (10% da nota final)

Proponha uma rede bayesiana para esse problema. Utilize o senso-comum, como relações causais esperadas, para propor uma topologia para a rede. No geral, você não vai ser avaliado pela topologia proposta, mas esse passo é importante para as demais etapas. Note que podemos descontar na nota final deste projeto caso a rede não tenha conexão alguma com a realidade.

Faça um programa que calcule as probabilidades marginais e condicionais a partir da base de dados **de treinamento** disponibilizada. As probabilidades a serem calculadas dependem da

arquitetura da rede. Como podem existir poucos dados para algumas combinações de valores, algumas normalizações não serão possíveis. Corrija este problema utilizando uma suavização aditiva (*Laplace Smoothing*) no cálculo das probabilidades empíricas. Por exemplo: assuma que todo evento acontece ao menos uma vez, e inicie as tabelas de contagem com um ao invés de zero. Pesquise sobre esse tema para entender as possíveis variações utilizadas na prática.

## Parte II – Classificação com rede bayesiana (20% da nota final)

Utilize a partição de teste da base para calcular a acurácia de classificação da rede: para cada exemplo de teste, calcule a classe mais provável entre as três classes existentes e compare a predição com a classe real existente na base de teste. Conte os acertos e a porcentagens de acertos entre todos os exemplos de teste.

Para saber se a sua rede bayesiana tem um bom desempenho, compare seu desempenho com algum classificador estado-da-arte. Um classificador simples e presente em diversos softwares de Aprendizado de Máquina são as Florestas Aleatórias (*Random Forest*). **Não é necessário** atingir um desempenho superior ou similar ao método utilizado na comparação, para obter nota máxima neste item.

## Parte III – Inferência da rede bayesiana (20% da nota final)

Implemente o algoritmo de inferência *likelihood weighting* para redes bayesianas. A implementação pode ser específica para a rede elaborada no item anterior. Alimente o algoritmo com as tabelas de probabilidades calculadas no item anterior. Esse algoritmo irá permitir que você faça inferências, mesmo quando não existem evidências para todos os atributos.

Utilize a rede bayesiana e o algoritmo de inferência para identificar fatores mais importantes que levam aos pacientes a utilizar ou não métodos contraceptivos, e relacione o tipo de uso de contraceptivos ao número de filhos. Utilize a rede bayesiana para realizar consultas do tipo “e se” contrastando fatores a favor (por exemplo, educação) e contra (por exemplo, religião). Efetue outras consultas para mostrar relações não óbvias e/ou interessante entre as diferentes variáveis (exemplos: relação de idade com o número de filhos, e relação de padrão de vida com exposição à mídia).

## Parte IV – Relatório (50% da nota final)

Escreva um relatório de até 10 páginas explicando as suas decisões de projeto, relatando as dificuldades encontradas e comparando as expectativas prévias com os resultados obtidos. Faça uma apresentação dos resultados obtidos por meio de gráficos, tabelas e matrizes de confusão.