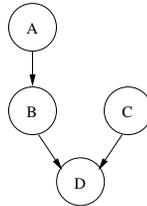


Taller 2

Entrega: Miércoles 7 de Noviembre de 2007 en clase

Minería de Datos - 2007-II

1. (30 %) Sean A , B , C y D variables binarias relacionadas de acuerdo con la siguiente red de creencia Bayesiana:



- a) Establezca las probabilidades condicionales y absolutas necesarias para especificar totalmente la red basándose en los siguientes ejemplos:

A	B	C	D
1	0	0	1
1	1	0	1
1	0	1	1
0	1	0	0
0	1	1	0
1	0	1	0
0	1	1	1
0	0	0	1
0	0	1	0

- b) Calcule la siguiente probabilidad condicional $P(A = 0|C = 1, D = 0)$
c) Dada la siguiente evidencia $A = 0$ y $C = 1$, ¿qué valor le daría a D ?
2. (50 %) Aplicación de técnicas de minería

- a) Obtenga el archivo `credit-simple.arff`
b) Entrene un modelo Naive Bayes Simple:
- 1) Use el 50 % de los datos para entrenar y el 50 % para probar
 - 2) Entrene el modelo
 - 3) ¿qué tal son los resultados? ¿cuántos errores cometió?
 - 4) ¿puede interpretar el modelo generado?
 - 5) Clasifique los siguientes ejemplos:
 - 'no checking',12,'existing paid',5000,500<=X<1000,'male single'
 - 0<=X<200,6,'existing paid',10000,<100,'male mar/wid'
- c) Entrene un modelo de redes neuronales:

- 1) Aplique los mismos pasos del ítem anterior
 - 2) Pruebe diferentes arquitecturas de la red (una capa interna con 1, 3, 5 y 10 neuronas), analice los resultados.
- d) Entrene un modelo de árbol de clasificación:
- 1) Aplique los mismos pasos del ítem b.
 - 2) Extraiga una regla sencilla para clasificar buenos deudores. ¿cuántos deudores clasifica la regla en el conjunto de entrenamiento?
 - 3) Haga experimentos usando el 25 %, el 10 % y el 5 % de los datos para entrenar y los datos restantes para probar. ¿qué sucede?
- e) Entrene un modelo sensitivo al costo:
- 1) Use el 50 % de los datos para entrenar y el 50 % para probar
 - 2) Especifique el algoritmo *Metacost* con *Naive Bayes Simple*
 - 3) Especifique una matriz de costos adecuada
 - 4) ¿qué tal es el desempeño en términos de costo?
 - 5) Cambie el algoritmo a un árbol de decisión. ¿qué tal es el desempeño en términos de costo del nuevo modelo?
3. (20 %) Predicción de series de tiempo
- a) El archivo `sunspots.csv` contiene información sobre la proporción de manchas en el sol desde 1701 a 1979.
 - b) Procese el archivo para crear un archivo de entrenamiento (primeros 200 datos) para una red neuronal que use una ventana de 10 valores hacia atrás.
 - c) Cree un archivo de prueba con los datos restantes.
 - d) Cargue los archivos de entrenamiento y de pruebas.
 - e) Entrene una red neuronal.
 - f) ¿qué tal es la predicción de la red neuronal?
 - g) Grafique los datos reales y la predicción.