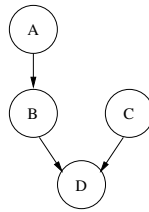


Taller 2: Aprendizaje Supervisado

Entrega: Miércoles 22 de Octubre de 2008 en clase

Minería de Datos - 2008-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 clasificación
- a) Obtenga el conjunto de datos `credit-g.arff` (<http://www.cs.umb.edu/~rickb/files/UCI/credit-g.arff>)
- b) Describa el conjunto de datos
- c) 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:

- '<0', 6, 'critical/other existing credit', radio/tv, 1169, 'no known savings', '>=7', 4, 'female div/dep/mar', none, 4, 'real estate', 67, none, own, 2, skilled, 1, yes, yes
- '>=200', 10, 'existing paid', 'domestic appliance', 1225, '<100', '<1', 3, 'female div/dep/mar', none, 1, car, 25, none, rent, 1, skilled, 1, none, yes

d) 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.

e) 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?

f) 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.