電子商務技術 期中考

- *** 請按照題號順序作答,不會的題目也請寫上題號 所有的計算過程都應寫出,否則扣分
- 1. 下圖為資料集與利用 NaiveBayes 訓練 watch-promo 後的預測結果,請依序回答以下問題:

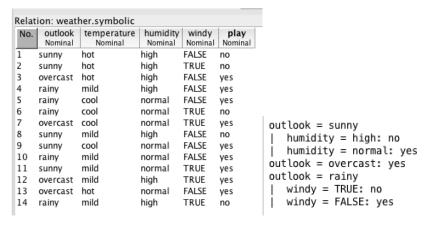
No.	Income-Range Nominal	Watch-Promo Nominal	Credit-Card-Ins Nominal	Sex Nominal	Age Numeric
1	40-50000	No	No	Male	45.0
2	30-40000	Yes	No	Female	40.0
3	40-50000	No	No	Male	42.0
4	30-40000	Yes	Yes	Male	43.0
5	50-60000	No	No	Female	38.0
6	20-30000	No	No	Female	55.0
7	30-40000	No	Yes	Male	35.0
8	20-30000	Yes	No	Male	27.0
9	30-40000	No	No	Male	43.0
10	30-40000	Yes	No	Female	41.0
11	40-50000	Yes	No	Female	43.0
12	20-30000	Yes	No	Male	29.0
13	50-60000	Yes	No	Female	39.0
14	40-50000	Yes	No	Male	55.0
15	20-30000	No	Yes	Female	19.0

=== Predictions on training set ===

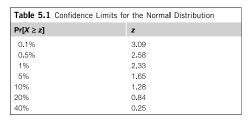
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- (1.1) Confusion matrix? (5%)
- (1.2) TP rate of class Yes? (5%)
- (1.3) Precision of class Yes? (5%)
- (1.4) Kappa statistic? (5%)
- (1.5) NaïveBayes classifier? (10%)
- (1.6) 依據上表詳細列出預測 instance 3 的演算過程。(5%)

- 2. 承上題資料集, instance 1-12 為訓練資料, 利用 IB1 預測 instance 13-15 的 watch-promo 值。請詳列計算過程。(15%)
- 3. 右下圖是 weather 測試集使用 ID3 後產生的結果,試據此回答以下問題:



- (3.1) 請繪出 Decision Tree,並標註葉節點(leaf)上的 instance 個數。 (5%)
- (3.2) 為何根節點 (root) 為 outlook ? (15%)
- 4. 某 J48 訓練出來的預測模型於 1000 筆資料的測試後得到 75%的正確率,試問它在 80%信心度條件下的真實正確率為何?(10%)



- 5. 底下是某資料集經 NaiveBayes 分類處理後的 Cost/Benefit Analysis 圖,請依此計算:
 - (5.1) Benefit? (5%)
 - (5.2) Random? (5%)
 - (5.3) % of Target? (5%)
 - (5.4) Lift? (5%)

