

2020 ECT 作業一

Mushrooms 資料集各個欄位的內容資訊位於最下方,提供參考

1. 用 Weka 軟體對 mushrooms.arff 利用 Naïve Bayes 進行 Supervised learning,選擇 “Use training set” , 設定 Attribute: type 為 Output, 在過程中對重要步驟截圖加以說明, 並回答以下問題：

- (a) 解釋 Classifier Output, Test data 的錯誤率為多少？有多少 Test dataset instances 被分類到有毒的(poisonous)但實際上屬於可食用的(edible)?請利用 Confusion matrix 解釋。(25%)
- (b) 在 Output predictions 結果中 “+” 代表的意義為何，請截圖並解釋。(10%)
- (c) 請使用 Visualize Classifier Errors, 解釋此圖與 Confusion matrix 之間的關係。(15%)

2. 用 python 對 mushrooms.csv 進行 Supervised learning 中的 Naïve Bayes 分析,並回答以下問題：

- (a) 在過程中對所有重要程式步驟進行截圖並加以說明，越詳盡越好。(15%)
- (b) 請問 mushrooms 資料集中共有多少 instance?是否包含空值的欄位(null)?(10%)
- (c) 請問欄位 “stalk_color_above_ring” 有幾種不同的 value?(5%)

(d) 請利用 `metrics.confusion_matrix ()` 呈現出混淆矩陣，並截圖加以

說明。(10%)

(e) 請利用 `metrics.classification_report()` 列出模型的準確率,並與

Weka 的結果比較何者較高?(10%)

- 繳交期限: 3/25(三) 中午 12:00
- 請轉檔為 PDF 格式, 檔名為: ECT_HW1_學號_版本.pdf, 並同時附上 python 的 ipynb 檔, 命名格式同上。
- 上傳至 LMS 作業區, 遲交一天扣該次作業成績 5%
- 補交請上傳至 HW1 補交區, 僅開放到 4/1(三) 中午 12:00

Attribute Information: (type edible=e, poisonous=p)

- cap-shape:
 - bell=b,
 - conical=c,
 - convex=x,
 - flat=f,
 - knobbed=k,
 - sunken=s
- cap-surface:
 - fibrous=f,
 - grooves=g,
 - scaly=y,
 - smooth=s
- cap-color:
 - brown=n,
 - buff=b,
 - cinnamon=c,
 - gray=g,
 - green=r,
 - pink=p,
 - purple=u,
 - red=e,
 - white=w,
 - yellow=y
- odor:
 - almond=a,
 - anise=l,
 - creosote=c,
 - fishy=y,
 - foul=f,
 - musty=m,
 - none=n,
 - pungent=p,
 - spicy=s
- stalk-shape:
 - enlarging=e,
 - tapering=t
- stalk-color-above-ring:

- brown=n,
- buff=b,
- cinnamon=c,
- gray=g,
- orange=o,
- pink=p,
- red=e,
- white=w,
- yellow=y

- stalk-color-below-ring:

- brown=n,
- buff=b,
- cinnamon=c
- ,gray=g,
- orange=o,
- pink=p
- ,red=e,
- white=w,
- yellow=y

- ring-number:

- none=n,
- one=o,
- two=t

- ring-type:

- cobwebby=c,
- evanescent=e,
- flaring=f,
- large=l,
- none=n,
- pendant=p,
- sheathing=s,
- zone=z

- population:

- abundant=a,
- clustered=c,
- numerous=n,
- scattered=s,
- several=v,

- solitary=y

• habitat:

- grasses=g,
- leaves=l,
- meadows=m,
- paths=p,
- urban=u,
- waste=w,
- woods=d