תרגיל מס' 1 לבית (אפשר להתחיל בכיתה)

היכרות עם WEKA

התרגיל להגשה לתאריך ה – 4.5.2016 שעה 23:59.59 יש להגיש באמצעות האתר

חלק א'

1. פתחו את הקובץ המוגדר בבסיס הנתונים של WEKA Labor.arff.
2. מהי התפלגות המחלקות (כמה אחוזים של מופעים ישנם מכל מחלקה ?
3. עבור כל Feature רשמו מהם ערכי המינימום, מקסימום, ממוצע, וסטיית התקן.

חלק ב'

הפעילו את אלגוריתם עץ ההחלטה J48 על הdataset המוגדר בקובץ Labor.arff.

1. עבור איזה צירוף ערכי פרמטרים (אם בכלל), יווצר עץ מינימאלי (עם שני עלים בלבד ?), מהי שגיאת המסווג במקרה זה ?
2. עבור איזה צירוף ערכי פרמטרים יווצר עץ "מקסימלי" (שלם ובעומק הגדול משניים) ? מהי שגיאת המסווג במקרה זה ?

הערה: את העץ ניתן לראות על ידי לחיצה על מקש עכבר ימני על הטקסט המציין את הניסוי

חלק ג'

חזרו על חלק ב' רק הפעם עם קובץ הנתונים contact-lenses.arff.

1. עבור איזה צירוף ערכי פרמטרים (אם בכלל), יווצר עץ מינימאלי (עם שני עלים בלבד ?), מהי שגיאת המסווג במקרה זה ?
2. עבור איזה צירוף ערכי פרמטרים יווצר עץ "מקסימלי" (שלם ובעומק הגדול משניים) ? מהי שגיאת המסווג במקרה זה ?

הערה: את העץ ניתן לראות על ידי לחיצה על מקש עכבר ימני על הטקסט המציין את הניסוי

חלק ד'

לפניכם קובץ נתונים של שבאמצעותו אופטיקאי יכול להחליט על המרשם להרכבת עדשות מגע לפציינטים. מדובר בעדשות רכות (Soft), קשות (hard), או ללא צורך בעדשות כלל (None). עליך לבנות תוכנית בשפת R שבאמצעותה תחלק את הקובץ ל TRAIN ו TEST, לאחר מכן יש להריץ מודל סיווג מסוג LOGISTIC REGRESSION ועץ החלטה J48, לכל אחד מהמודלים יש להציג את ה – CONFUSION MATRIX, ולרשום מה אחוז הדיוק של המודל. (עליך לייצר חלוקה של 30% לTEST ו 70% לTRAIN). יש לצרף את הקוד לקובץ ההגשה. באם קיבלת הבדל בתוצאות הדיוק של 2 המודלים – נא ציין ממה לדעתך נובע הבדל זה ?

קובץ הנתונים:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Age** | **Spectacle-prescrip** | **Astigmatism** | **Tear-prod-rate** | **Contact-lenses** |
| young | myope | no | normal | soft |
| young | myope | yes | reduced | none |
| young | myope | yes | normal | hard |
| young | hypermetrope | no | reduced | none |
| young | hypermetrope | no | normal | soft |
| young | hypermetrope | yes | reduced | none |
| pre-presbyopic | myope | no | reduced | none |
| pre-presbyopic | myope | no | normal | soft |
| pre-presbyopic | myope | yes | normal | hard |
| pre-presbyopic | hypermetrope | no | reduced | none |
| pre-presbyopic | hypermetrope | no | normal | soft |
| pre-presbyopic | hypermetrope | yes | reduced | none |
| pre-presbyopic | hypermetrope | yes | normal | none |
| presbyopic | myope | no | normal | none |
| presbyopic | myope | yes | reduced | none |
| presbyopic | myope | yes | normal | hard |
| presbyopic | hypermetrope | no | reduced | none |
| presbyopic | hypermetrope | no | normal | soft |
| presbyopic | hypermetrope | yes | reduced | none |
| presbyopic | hypermetrope | yes | normal | none |