Homework Machine Learning 3cw1

205Q001019

ທ້າວ ນູຊົ່ວ ເຮີ

ແກ້ບົດເຝິກຫັດ ຂໍ້ 2.2

5. ຈົ່ງນຳໃຊ້ຄຳສັ່ງ Python ເພື່ອແບ່ງຊຸດຂໍ້ມູນ Data.csvອອກເປັນ 2 ຊຸດຄື: X (Features) ແລະ y (Label)

ຄຳສັ່ງຕົວຢ່າງ:

df= pd.read\_csv('Data.csv')

X = df.iloc[:, :-1].values

y = df.iloc[:, -1].values

X:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Country | Age | Salary |
| 0 | France | 44.0 | 72000.0 |
| 1 | Spain | 27.0 | 48000.0 |
| 2 | Germany | 30.0 | 54000.0 |
| 3 | Spain | 38.0 | 61000.0 |
| 4 | Germany | 40.0 | nan |
| 5 | France | 35.0 | 58000.0 |
| 6 | Spain | nan | 52000.0 |
| 7 | France | 48.0 | 79000.0 |
| 8 | Germany | 50.0 | 83000.0 |
| 9 | France | 37.0 | 67000.0 |

y (Label)

|  |  |
| --- | --- |
|  | Purscased |
| 0 | No |
| 1 | yes |
| 2 | no |
| 3 | no |
| 4 | yes |
| 5 | yes |
| 6 | no |
| 7 | yes |
| 8 | No |
| 9 | yes |

6. ຈົ່ງນຳໃຊ້ຄຳສັ່ງ Python ເພື່ອແປງຄ່າຂອງ X ແລະ y

ຄຳສັ່ງຕົວຢ່າງ:

from sklearn.compose import ColumnTransformer

from sklearn.preprocessing import OneHotEncoder

ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [0])], remainder='passthrough')

X = np.array(ct.fit\_transform(X))

ຄ່າຂອງX ທີ່ແປງແລ້ວຮອບ 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | Country | Age | Salary |
| 0 | 1.0 0.0 0.0 | 44.0 | 72000.0 |
| 1 | 0.0 0.0 1.0 | 27.0 | 48000.0 |
| 2 | 0.0 1.0 0.0 | 30.0 | 54000.0 |
| 3 | 0.0 0.0 1.0 | 38.0 | 61000.0 |
| 4 | 0.0 1.0 0.0 | 40.0 | 63777.77777777778 |
| 5 | 1.0 0.0 0.0 | 35.0 | 58000.0 |
| 6 | 0.0 0.0 1.0 | 38.77777777777778 | 52000.0 |
| 7 | 0.0 1.0 0.0 | 48.0 | 79000.0 |
| 8 | 0.0 1.0 0.0 | 50.0 | 83000.0 |
| 9 | 1.0 0.0 0.0 | 37.0 | 67000.0 |

ຄຳສັ່ງຕົວຢ່າງ:

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()

X\_train[:, 3:] = sc.fit\_transform(X\_train[:, 3:])

X\_test[:, 3:] = sc.transform(X\_test[:, 3:])

ຄ່າຂອງX ທີ່ແປງແລ້ວຮອບ 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | Country | Age | Salary |
| 0 | 1.0 0.0 0.0 | -0.19159184384578545 | -1.0781259408412425 |
| 1 | 0.0 1.0 0.0 | -0.014117293757057777 | -0.07013167641635372 |
| 2 | 1.0 0.0 0.0 | 0.566708506533324 | 0.633562432710455 |
| 3 | 0.0 0.0 1.0 | -0.30453019390224867 | -0.30786617274297867 |
| 4 | 0.0 0.0 1.0 | -1.9018011447007988 | -1.420463615551582 |
| 5 | 1.0 0.0 0.0 | 1.1475343068237058 | 1.232653363453549 |
| 6 | 0.0 1.0 0.0 | 1.4379472069688968 | 1.5749910381638885 |
| 7 | 1.0 0.0 0.0 | -0.7401495441200351 | -0.5646194287757332 |
| 8 | 0.0 1.0 0.0 | -1.4661817944830124 | -0.9069571034860727 |
| 9 | 1.0 0.0 0.0 | -0.44973664397484414 | 0.2056403393225306 |

ຄຳສັ່ງຕົວຢ່າງ:

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

y = le.fit\_transform(y)

ຄ່າຂອງ yແປງແລ້ວ

|  |  |
| --- | --- |
|  | Purscased |
| 0 | 0 |
| 1 | 1 |
| 2 | 0 |
| 3 | 0 |
| 4 | 1 |
| 5 | 1 |
| 6 | 0 |
| 7 | 1 |
| 8 | 0 |
| 9 | 1 |