DT 1

December 18, 2022

[45]:
```
import pandas as pd
```

[46]:
```
df = pd.read_csv("salaries.csv")
df.head()
```

[46]:

	company	job	degree	salary_more_then_100k
0	google	sales executive	bachelors	0
1	google	sales executive	masters	0
2	google	business manager	bachelors	1
3	google	business manager	masters	1
4	google	computer programmer	bachelors	0

[47]:
```
inputs = df.drop('salary_more_then_100k',axis='columns')
```

[48]:
```
target = df['salary_more_then_100k']
```

[49]:
```
from sklearn.preprocessing import LabelEncoder
le_company = LabelEncoder()
le_job = LabelEncoder()
le_degree = LabelEncoder()
```

[50]:
```
inputs['company_n'] = le_company.fit_transform(inputs['company'])
inputs['job_n'] = le_job.fit_transform(inputs['job'])
inputs['degree_n'] = le_degree.fit_transform(inputs['degree'])
```

[51]:
```
inputs
```

[51]:

	company	job	degree	company_n	job_n	degree_n
0	google	sales executive	bachelors	2	2	0
1	google	sales executive	masters	2	2	1
2	google	business manager	bachelors	2	0	0
3	google	business manager	masters	2	0	1
4	google	computer programmer	bachelors	2	1	0
5	google	computer programmer	masters	2	1	1
6	abc pharma	sales executive	masters	0	2	1
7	abc pharma	computer programmer	bachelors	0	1	0
8	abc pharma	business manager	bachelors	0	0	0
9	abc pharma	business manager	masters	0	0	1

```
10    facebook      sales executive  bachelors          1      2         0
11    facebook      sales executive    masters          1      2         1
12    facebook     business manager  bachelors          1      0         0
13    facebook     business manager    masters          1      0         1
14    facebook  computer programmer  bachelors          1      1         0
15    facebook  computer programmer    masters          1      1         1
```

[52]:
```
inputs_n = inputs.drop(['company','job','degree'],axis='columns')
```

[53]:
```
inputs_n
```

[53]:
```
    company_n  job_n  degree_n
0           2      2         0
1           2      2         1
2           2      0         0
3           2      0         1
4           2      1         0
5           2      1         1
6           0      2         1
7           0      1         0
8           0      0         0
9           0      0         1
10          1      2         0
11          1      2         1
12          1      0         0
13          1      0         1
14          1      1         0
15          1      1         1
```

[54]:
```
target
```

[54]:
```
0     0
1     0
2     1
3     1
4     0
5     1
6     0
7     0
8     0
9     1
10    1
11    1
12    1
13    1
14    1
15    1
```

```
Name: salary_more_then_100k, dtype: int64
```

[55]:
```
from sklearn import tree
model = tree.DecisionTreeClassifier()
```

[56]:
```
model.fit(inputs_n, target)
```

[56]:
```
DecisionTreeClassifier()
```

[57]:
```
model.score(inputs_n,target)
```

[57]:
```
1.0
```

[58]:
```
model.predict([[2,1,0]])
```

```
C:\Users\Deepak\ana-conda-3\lib\site-packages\sklearn\base.py:450: UserWarning:
X does not have valid feature names, but DecisionTreeClassifier was fitted with
feature names
  warnings.warn(
```

[58]:
```
array([0], dtype=int64)
```

[59]:
```
model.predict([[2,1,1]])
```

```
C:\Users\Deepak\ana-conda-3\lib\site-packages\sklearn\base.py:450: UserWarning:
X does not have valid feature names, but DecisionTreeClassifier was fitted with
feature names
  warnings.warn(
```

[59]:
```
array([1], dtype=int64)
```

[ ]:

[ ]:

[ ]: