ML 9 - Nominal or OHE (One Hot Encoding) By Virat Tiwari

December 1, 2023

1 Nominal / OHE (One Hot Encoding)

```python
[1]: # We import pandas for creating the dataframe
     # We use sklearn for importing the OneHotEncoder

     import pandas as pd
     from sklearn.preprocessing import OneHotEncoder
```

```python
[4]: # Here we create a simple Dataframe
     # color is our feature in dataframe
     # Now we create a sample dataframme with a cateorical variable
     # pd.DataFrame ( { } ) function is used for making dataframe

     df=pd.DataFrame({
         "color":["red","blue","green","green","red","blue"]
     })
```

```python
[5]: # This is our dataframe

     df
```

```
[5]:    color
     0    red
     1   blue
     2  green
     3  green
     4    red
     5   blue
```

```python
[12]: # Create an instance of onehotencoder
      # This instance is used for the fir transform operation that used for␣
      ↪converting the datatype from categorical to numerical

      encoder=OneHotEncoder()
```

```python
[13]: # encoder.fit_transform ( ) is used for tranforming the data type from␣
      ↪categorical to numerical

      encoded=encoder.fit_transform(df[["color"]])
```

```python
[20]: # pd.DataFrame(encoded.toarray(),columns=encoder.get_feature_names_out()) is␣
      ↪used for converting the dxata and this is the last step as well

      import pandas as pd
      encoded_df=pd.DataFrame(encoded.toarray(),columns=encoder.
      ↪get_feature_names_out())
```

```python
[21]: # Through we get the array

      encoder.get_feature_names_out()
```

```
[21]: array(['color_blue', 'color_green', 'color_red'], dtype=object)
```

```python
[22]: # So this is our final ouput that shows the coversion

      encoded_df
```

```
[22]:    color_blue  color_green  color_red
     0         0.0          0.0        1.0
     1         1.0          0.0        0.0
     2         0.0          1.0        0.0
     3         0.0          1.0        0.0
     4         0.0          0.0        1.0
     5         1.0          0.0        0.0
```

```python
[23]: # concat  ( ) function is used for adding the data like here we add previous␣
      ↪and current converted data

      pd.concat([df,encoded_df],axis=1)
```

```
[23]:    color  color_blue  color_green  color_red
     0    red         0.0          0.0        1.0
     1   blue         1.0          0.0        0.0
     2  green         0.0          1.0        0.0
     3  green         0.0          1.0        0.0
     4    red         0.0          0.0        1.0
     5   blue         1.0          0.0        0.0
```

P R A C T I S E

```python
[25]: import seaborn as sns
```

```
[27]: df=sns.load_dataset("tips")
```

```
[28]: df.head(
```

```
[28]:    total_bill   tip     sex smoker  day    time  size
      0        16.99  1.01  Female     No  Sun  Dinner     2
      1        10.34  1.66    Male     No  Sun  Dinner     3
      2        21.01  3.50    Male     No  Sun  Dinner     3
      3        23.68  3.31    Male     No  Sun  Dinner     2
      4        24.59  3.61  Female     No  Sun  Dinner     4
```

```
[38]: df1=pd.DataFrame(df)
```

```
[50]: encoder=OneHotEncoder()
```

```
[51]: encoded=encoder.fit_transform(df1[["sex"]])
```

```
[52]: import pandas as pd
      encoded_df1=pd.DataFrame(encoded.toarray(),columns=encoder.
       ↪get_feature_names_out())
```

```
[53]: encoder.get_feature_names_out()
```

```
[53]: array(['sex_Female', 'sex_Male'], dtype=object)
```

```
[56]: encoded_df1.head()
```

```
[56]:    sex_Female  sex_Male
      0         1.0       0.0
      1         0.0       1.0
      2         0.0       1.0
      3         0.0       1.0
      4         1.0       0.0
```

```
[58]: pd.concat([df,encoded_df1],axis=1).head()
```

```
[58]:    total_bill   tip     sex smoker  day    time  size  sex_Female  sex_Male
      0        16.99  1.01  Female     No  Sun  Dinner     2         1.0       0.0
      1        10.34  1.66    Male     No  Sun  Dinner     3         0.0       1.0
      2        21.01  3.50    Male     No  Sun  Dinner     3         0.0       1.0
      3        23.68  3.31    Male     No  Sun  Dinner     2         0.0       1.0
      4        24.59  3.61  Female     No  Sun  Dinner     4         1.0       0.0
```

THANK YOU SO MUCH !!

YOURS VIRAT TIWARI :)