

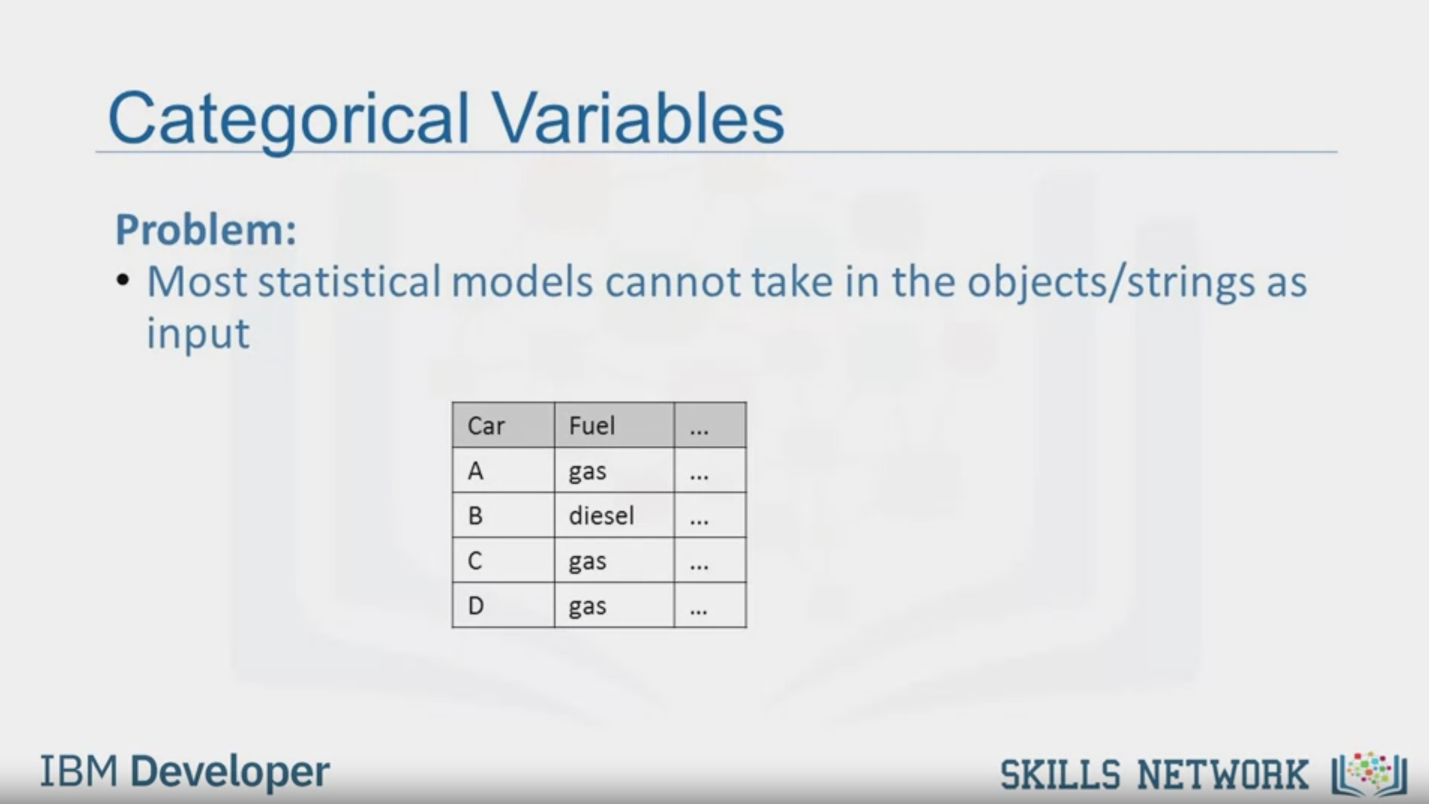
Turning categorical variables

into quantitative variables

in Python

IBM Developer

SKILLS NETWORK



Categorical Variables

Problem:

• Most statistical models cannot take in the objects/strings as

input

Car

A

B

C

D

Fuel

gas

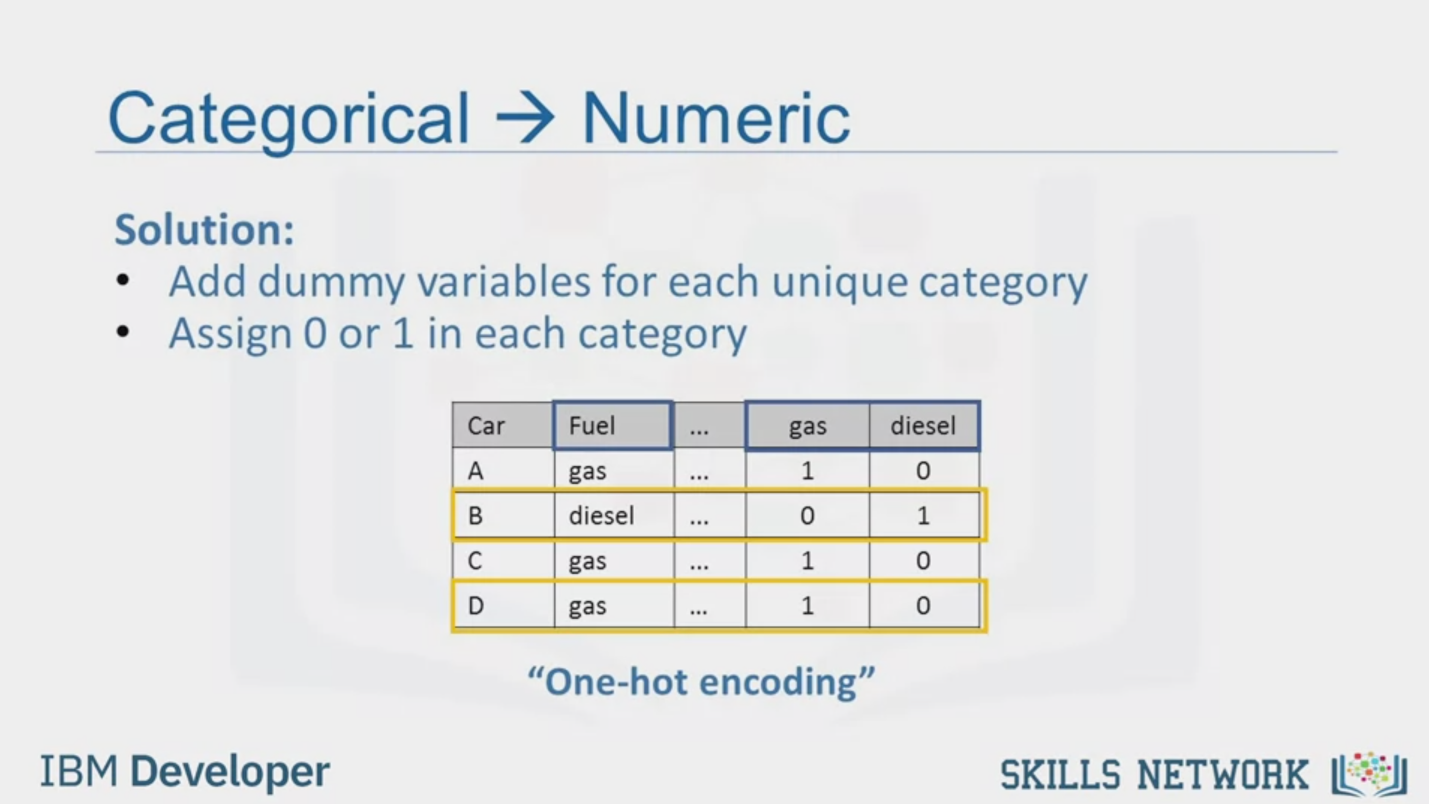
diesel

gas

gas

IBM Developer

SKILLS NETWORK



Categorical > Numeric

Solution:

Add dummy variables for each unique category

Assign 0 or 1 in each category

Car

A

B

C

D

Fuel

gas

diesel

gas

gas

gas

1

0

1

1

diesel

0

1

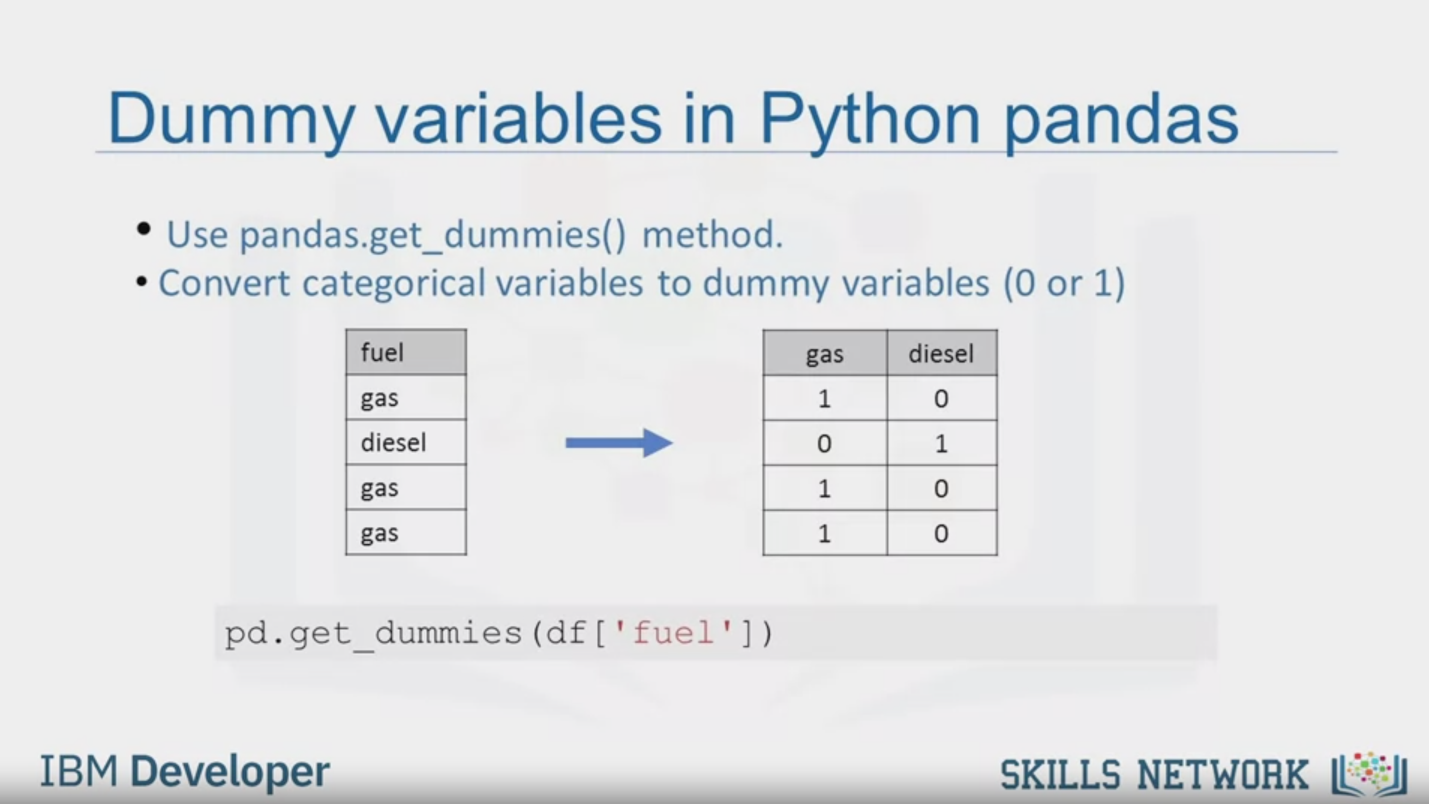
0

0

"One-hot encoding"

IBM Developer

SKILLS NETWORK



Dummy variables in Python pandas

• Use pandas.get\_dummies() method.

• Convert categorical variables to dummy variables (0 or 1)

fuel

gas

diesel

gas

gas

gas

1

0

1

1

diesel

0

1

0

0

pd.get\_dummies (df [ 'fuel'])

IBM Developer

SKILLS NETWORK



SKILLS NETWORK

IBM Developer

In this video, we'll discuss how to turn

categorical variables into quantitative variables in Python.

Most statistical models cannot take in objects or strings as input

and for model training only take the numbers as inputs.

In the car data set,

the fuel type feature as a categorical variable has two values,

gas or diesel, which are in string format.

For further analysis, Jerry has to convert

these variables into some form of numeric format.

We encode the values by adding new features corresponding to

each unique element in the original feature we would like to encode.

In the case where the feature fuel has two unique values,

gas and diesel, we create two new features, gas and diesel.

When a value occurs in the original feature,

we set the corresponding value to one in the new feature.

The rest of the features are set to zero.

In the fuel example,

for car B, the fuel value is diesel.

Therefore we set the feature diesel equal to one and the gas feature to zero.

Similarly, for Car D,

the fuel value is gas.

Therefore we set the feature gas equal to one and the feature diesel equal to zero.

This technique is often called "One-hot encoding.

In Pandas, we can use get\_dummies

method to convert categorical variables to dummy variables.

In Python, transforming categorical variables to dummy variables is simple.

Following the example, pd.get\_dummies method gets

the fuel type column and creates the data frame dummy\_variable\_1.

The get\_dummies method automatically generates a list of numbers,

each one corresponding to a particular category of the variable.