Paul Polsinelli

IFT 410 - 22485

4/11/2021

# Lab 11: Transforming the Mushroom Dataset

**Part III: Writing the Transformation Script**

import pandas as pd

import sklearn.datasets as sd

df = pd.read\_csv("agaricus-lepiota.csv", header=None)

print(df)

class\_labels = df[0].unique()

print("The class labels are:", class\_labels)

CL = {class\_labels[0]: '0', class\_labels[1]: '1'}

new\_class\_labels = df[0].map(CL)

print(new\_class\_labels)

new\_df = pd.DataFrame()

new\_df[0] = new\_class\_labels

col= 1

m = len(df.columns)

for i in range(1, m):

if i != 11:

x = pd.get\_dummies(df[i])

for j in range(0, len(x.columns)):

new\_df[col] = x[ x.columns[j] ]

col += 1

print(new\_df)

class\_labels = new\_df[0]

train\_data = new\_df.iloc[:, 1:len(new\_df.columns)]

sd.dump\_svmlight\_file(train\_data, class\_labels, "mushroom\_prepared.libsvm", False)



