Real_data_example_Binomial-MGWR

May 18, 2020

Notebook Outline:

0.0.1 Set up Cells

```
In [1]: import sys
        sys.path.append("C:/Users/msachde1/Downloads/Research/Development/mgwr")

In [2]: import warnings
        warnings.filterwarnings("ignore")
        import pandas as pd
        import numpy as np

        from mgwr.gwr import GWR
        from spglm.family import Gaussian, Binomial, Poisson
        from mgwr.gwr import MGWR
        from mgwr.sel_bw import Sel_BW
        import multiprocessing as mp
        pool = mp.Pool()
        from scipy import linalg
        import numpy.linalg as la
        from scipy import sparse as sp
        from scipy.sparse import linalg as spla
        from spreg.utils import spdot, spmultiply
        from scipy import special
        import libpysal as ps
```

```
import seaborn as sns
import matplotlib.pyplot as plt
from copy import deepcopy
import copy
from collections import namedtuple
import spglm
```

0.0.2 Clearwater Landslides Dataset

Clearwater data - downloaded from link: https://sgsup.asu.edu/sparc/multiscale-gwr

In [3]:
```
data_p = pd.read_csv("C:/Users/msachde1/Downloads/logistic_mgwr_data/landslides.csv")
```

In [4]:
```
data_p.head()
```

Out[4]:

	UserID	X	Y	Elev	Slope	SinAspct	CosAspct	\
0	1	616168.5625	5201076.5	1450.475	27.44172	0.409126	-0.912478	
1	2	624923.8125	5201008.5	1567.476	21.88343	-0.919245	-0.393685	
2	3	615672.0000	5199187.5	1515.065	38.81030	-0.535024	-0.844837	
3	4	615209.3125	5199112.0	1459.827	26.71631	-0.828548	-0.559918	
4	5	616354.6875	5198945.5	1379.442	27.55271	-0.872281	-0.489005	

	AbsSouth	Landslid	DistStrm
0	24.1499	1	8.506
1	66.8160	1	15.561
2	32.3455	1	41.238
3	55.9499	1	17.539
4	60.7248	1	35.023

0.0.3 Univariate example

GWR Binomial model with independent variable, x = slope

In [5]:
```
coords = list(zip(data_p['X'],data_p['Y']))
y = np.array(data_p['Landslid']).reshape((-1,1))
elev = np.array(data_p['Elev']).reshape((-1,1))
slope = np.array(data_p['Slope']).reshape((-1,1))
SinAspct = np.array(data_p['SinAspct']).reshape(-1,1)
CosAspct = np.array(data_p['CosAspct']).reshape(-1,1)
X = np.hstack([elev,slope,SinAspct,CosAspct])
x = CosAspct

X_std = (X-X.mean(axis=0))/X.std(axis=0)
x_std = (x-x.mean(axis=0))/x.std(axis=0)
y_std = (y-y.mean(axis=0))/y.std(axis=0)
```

In [6]:
```
bw=Sel_BW(coords,y,x_std,family=Binomial(),constant=False).search()
gwr_mod=GWR(coords,y,x_std,bw=bw,family=Binomial(),constant=False).fit()
bw
```

Out[6]:
```
108.0
```

Running the function with family = Binomial()

Bandwidths check

```
In [7]: selector = Sel_BW(coords,y,x_std,family=Binomial(),multi=True,constant=False)
        selector.search(verbose=True)

Current iteration: 1 ,SOC: 0.0752521
Bandwidths: 108.0
Current iteration: 2 ,SOC: 0.0213201
Bandwidths: 184.0
Current iteration: 3 ,SOC: 5.8e-05
Bandwidths: 184.0
Current iteration: 4 ,SOC: 1e-06
Bandwidths: 184.0


Out[7]: array([184.])

In [8]: mgwr_mod = MGWR(coords, y,x_std,selector,family=Binomial(),constant=False).fit()

HBox(children=(IntProgress(value=0, description='Inference', max=1), HTML(value='')))
```

Parameter check

```
In [9]: mgwr_mod.bic

Out[9]: 325.23949237389036

In [10]: gwr_mod.bic

Out[10]: 338.19722049287054
```

0.0.4 Multivariate example

```
In [12]: bw=Sel_BW(coords,y,X_std,family=Binomial(),constant=True).search()
         gwr_mod=GWR(coords,y,X_std,bw=bw,family=Binomial(),constant=True).fit()
         bw

Out[12]: 121.0
```

Bandwidth check

```
In [13]: selector = Sel_BW(coords,y,X_std,family=Binomial(),multi=True,constant=True)
         selector.search(verbose=True)
```

```
Current iteration: 1 ,SOC: 0.116124
Bandwidths: 43.0, 62.0, 191.0, 100.0, 108.0
Current iteration: 2 ,SOC: 0.0266811
Bandwidths: 43.0, 106.0, 210.0, 100.0, 184.0
Current iteration: 3 ,SOC: 0.0008147
Bandwidths: 43.0, 106.0, 210.0, 100.0, 184.0
Current iteration: 4 ,SOC: 5.28e-05
Bandwidths: 43.0, 106.0, 210.0, 100.0, 184.0
Current iteration: 5 ,SOC: 5.3e-06
Bandwidths: 43.0, 106.0, 210.0, 100.0, 184.0
```

```
Out[13]: array([ 43., 106., 210., 100., 184.])
```

```
In [14]: mgwr_mod = MGWR(coords, y,X_std,selector,family=Binomial(),constant=True).fit()
```

```
HBox(children=(IntProgress(value=0, description='Inference', max=1), HTML(value='')))
```

AIC, AICc, BIC check

```
In [15]: gwr_mod.aicc, mgwr_mod.aicc
```

```
Out[15]: (264.9819711678866, 251.85376815296377)
```

0.0.5 Global model check

```
In [16]: selector=Sel_BW(coords,y,X_std,multi=True,family=Binomial(),constant=True)
         selector.search(verbose=True,multi_bw_min=[239,239,239,239,239], multi_bw_max=[239,239
```

```
Current iteration: 1 ,SOC: 0.6120513
Bandwidths: 239.0, 239.0, 239.0, 239.0, 239.0
Current iteration: 2 ,SOC: 0.0594775
Bandwidths: 239.0, 239.0, 239.0, 239.0, 239.0
Current iteration: 3 ,SOC: 0.0025897
Bandwidths: 239.0, 239.0, 239.0, 239.0, 239.0
Current iteration: 4 ,SOC: 0.0001289
Bandwidths: 239.0, 239.0, 239.0, 239.0, 239.0
Current iteration: 5 ,SOC: 1.17e-05
Bandwidths: 239.0, 239.0, 239.0, 239.0, 239.0
Current iteration: 6 ,SOC: 1.2e-06
Bandwidths: 239.0, 239.0, 239.0, 239.0, 239.0
```

```
Out[16]: array([239., 239., 239., 239., 239.])

In [17]: mgwr_mod = MGWR(coords, y,X_std,selector,family=Binomial(),constant=True).fit()

HBox(children=(IntProgress(value=0, description='Inference', max=1), HTML(value='')))



In [18]: gwr_mod.summary()

===============================================================================
Model type                                                         Binomial
Number of observations:                                                 239
Number of covariates:                                                     5

Global Regression Results
-------------------------------------------------------------------------------
Deviance:                                                         266.246
Log-likelihood:                                                  -133.123
AIC:                                                              276.246
AICc:                                                             276.504
BIC:                                                            -1015.246
Percent deviance explained:                                         0.182
Adj. percent deviance explained:                                    0.168

Variable                              Est.         SE  t(Est/SE)    p-value
------------------------------ ---------- ---------- ---------- ----------
X0                                 0.389      0.150      2.591      0.010
X1                                -0.784      0.166     -4.715      0.000
X2                                 0.654      0.168      3.881      0.000
X3                                 0.039      0.149      0.264      0.792
X4                                -0.371      0.156     -2.381      0.017

Geographically Weighted Regression (GWR) Results
-------------------------------------------------------------------------------
Spatial kernel:                                     Adaptive bisquare
Bandwidth used:                                               121.000

Diagnostic information
-------------------------------------------------------------------------------
Effective number of parameters (trace(S)):                     23.263
Degree of freedom (n - trace(S)):                             215.737
Log-likelihood:                                              -106.599
AIC:                                                          259.725
AICc:                                                         264.982
BIC:                                                          340.598
Percent deviance explained:                                     0.345
```

```
Adjusted percent deviance explained:                                0.274
Adj. alpha (95%):                                                   0.011
Adj. critical t value (95%):                                        2.571

Summary Statistics For GWR Parameter Estimates
------------------------------------------------------------------------------
Variable                   Mean        STD        Min     Median        Max
-------------------- ---------- ---------- ---------- ---------- ----------
X0                        0.459      0.360     -0.360      0.436      1.232
X1                       -0.824      0.479     -2.128     -0.729     -0.095
X2                        0.567      0.390     -0.030      0.600      1.328
X3                        0.103      0.270     -0.473      0.183      0.565
X4                       -0.331      0.247     -1.118     -0.287      0.096
==============================================================================
```

```
In [19]: np.mean(mgwr_mod.params,axis=0)
```

```
Out[19]: array([ 0.19936242, -0.3251776 ,  0.32069312,  0.04295657, -0.20408904])
```

```
In [20]: mgwr_mod.bic, gwr_mod.bic
```

```
Out[20]: (303.9521120546862, 340.5982180538755)
```