

GLO-7050

Gabriel Rieger Junqueira

82 / 100

Été 2020

Toutes les formules que j'ai implémenté ont été pris dans Wikipedia.

Question 1

Partie sans régularisation

Pour l'entraînement ça va bien mais pour la validation le polynôme de degré 20 est trop complexe pour les données, ce qui donne un MSE élevé. -8% Nous espérons que vous justifiez que ce modèle est trop adapté ou non (overfitting or underfitting?)

Partie avec régularisation

Pour l'entraînement le MSE reste environ le même que sans régularisation et pour la validation

('indice du meilleur ajustement est', 1)

('La valeur de lambda est', 0.05)

Question 2

On a comme résultat

('MSE pour train', array([0.12725264]))

('MSE pour valid', array([0.12706128]))

('MSE pour test', array([0.09673425]))

En variant le taux d'apprentissage et le pas et l'époque on a:

('MSE pour Train', array([[28.15979119],
[27.88315889],
[26.54040931],
[20.73947091],
[6.11438077],
[0.15725863]]))

('MSE pour Valid', array([[28.02452976],
[27.75029372],
[26.41910702],
[20.66662939],
[6.14222917],
[0.16671466]]))

('MSE pour Test', array([[27.64924999],
[27.37723957],
[26.05693467],
[20.35316891],
[5.97628126],
[0.12624132]]))

Question 3

Pour ce cas remplacer pour la moyenne c'est assez bien, d'autres approches seraient remplacer pour la médiane ou la mode.

En théorie le meilleur remplacement sera par un modèle un suivant une loi de distribution normal ou la même distribution de chaque attribut, cependant cela ce montre impraticable dans la vraie vie, car on a 122 colonnes à modéliser.

Les erreurs sont:

pour chaque plies

```
('MSE train', array([0.01653138, 0.01584503, 0.0159012 , 0.0162319 , 0.01532885]))  
( 'MSE valid', array([0.0171729 , 0.02039048, 0.02132102, 0.01867044, 0.63520837]))
```

Max et Min

```
('MSE minimum train est:', 0.015328850722763472)
```

```
('MSE minimum valid est:', 0.017172903686204253)
```

Moyennée

```
('Moyenne pour MSE train:', 0.01596767260351408)  
( 'Moyenne pour MSE valid:', 0.14255264520197777)
```

Paramètres d'apprentissage avec la régression Ridge -10%

```
('Moyenne MSE train Ridge: ', array([6.97614122e-02, 2.75041234e+05, 1.10017819e+06,  
2.47541094e+06,  
4.40073948e+06, 6.87616381e+06, 9.90168393e+06, 1.34772998e+07,  
1.76030116e+07, 2.22788191e+07]))  
( 'Moyenne MSE valid Ridge: ', array([6.86016470e-02, 2.74993633e+05, 1.09998754e+06,  
2.47498179e+06,  
4.39997638e+06, 6.87497131e+06, 9.89996658e+06, 1.34749622e+07,  
1.75999582e+07, 2.22749545e+07]))
```

```
paramètre min', array([ 36.28443233, 156.23042373, 114.99641945, 201.26614128,  
53.80364959, 73.10535617, 147.23626139, 174.82544109,  
119.41929502, 150.05140521, 38.12877574, 248.78623914,  
93.39504436, 171.17193401, 89.21108127, 135.12391351,  
168.92485889, 151.60424111, 158.75747144, 99.64316457,  
97.49919885, 113.29952725, 83.3121471 , 64.64550216,  
99.9864204 , 88.30761116, 117.48171755, 39.34746982,  
143.02391039, 137.57860922, 163.94778896, 102.01592124,  
156.97197653, 152.00801002, 130.63244684, 147.84509388,  
152.24602131, 130.78325515, 189.1012454 , 166.1748574 ,  
198.68619694, 201.08332107, 173.73574721, 161.01027413,  
160.69871747, 177.53808423, 156.7684103 , 169.13278501,  
170.23781331, 30.08740265, 142.036183 , 19.07882121,  
120.9085225 , 137.70978969, 152.08043754, 164.64598001,  
79.17247185, 80.65298871, 82.91509879, 81.96181385,
```

```
248.76835519, 73.90817334, 116.42427379, 104.85440191,  
156.30070884, 162.15474167, 153.26431537, 158.15746554,  
94.92093464, 195.34226082, 77.73430044, 48.53161652,  
224.63279303, 159.85248018, 105.09301195, 149.62276245,  
160.08259678, 129.2329253 , 107.70022881, 73.99751029,  
75.2433647 , 78.28928912, 98.92035469, 110.16652974,  
125.28831186, 114.15927754, 185.58054577, 157.84034353,  
140.95821474, 23.7770013 , 19.30756263, 88.94285011,  
202.87858294, 174.63920861, 220.04472305, 221.22430348,  
28.5761567 , 1612.02741299, 312.61561513, 87.06641218,  
38.60427597, 77.19189255, 120.17203834, -1459.46449752,  
230.9417216 , 241.72030346, 83.43450696, 48.24012652,  
39.48705617, 95.23866943, 29.37665201, 191.18146046,  
104.68774486, 30.3728595 , 97.35230862, 66.05553967,  
62.27587762, 30.53486662, 237.12420676, 154.72003706,
```

```
print('Paramètre max',W_ridge_3[np.argmax(MSE_valid_ridge_3)])
```

index 48 is out of bounds for axis 0 with size 10

J'ai eu des problèmes pour le trouver pourquoi vous ne l'as pas trouvé? Avez-vous vérifié l'index dans chaque dimension?