Linear Regression with Newton Search

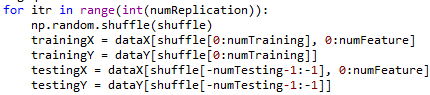
20150669 전예일

This code have main class, ‘RegressionAnalysis’ and its attributes and methods. This program, first,

open the file named day.csv. This program has two course, train, and test. Train course is finding better parameters which can predict Y more nearly using linear regression and Newton methods. In training course, it train parameters by linear regression and Newton methods, so when we do more train course, parameter will find answer more exactly. Test course just calculate MSE.

프로그램은 file을 읽은 뒤

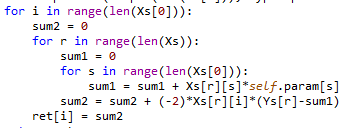
 에서 10, 20, … 100 까지 증가하는 numTraining의 값으로

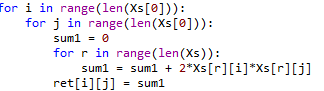


위의 과정을 수행한다. (numReplication으로 수차례 시행하고 그 평균값을 구해줌으로써 좀 더 정확한 값을 찾는다) 여기서 Training 의 경우 shuffle된 일정구간을 train시켜 better parameter 을 찾아내기에 상대적으로 작은 MSE를 좀더 빨리 찾아내겠지만 Testing의 경우

 처럼 고정된 numTesting의 값에서 shuffle[-51:-1] 까지의 범위의 MSE를 return 하기 때문에 training의 경우보다 MSE가 느리게 0에 수렴하게 된다.

Homework의 introduction에 나온 것처럼  에 따라 parameter을 갱신하기 때문에 이와 같이 parameter을 갱신하여 준다. 이를위하여 firstDeriv와 invHessian을 구해야 하고 introduction에 나온 식을 이용하여 다음과 같이 firstDeriv를 구해주고,

 ( i는 Xs의 가로길이에 해당하는 len(Xs[0])만큼, r은 세로길이에 해당하는 len(Xs)만큼 반복하여 값을 구해줌)

이와 같이 SecondeDeriv 도 구해준다..

( 이것과 이것이 같은 식임을 알 수 있다.)

MeansquareError의 경우 Y의 값과 Y hat의 값의 차를 제곱한 것의 평균이기에

를 통해 Y와 Y hat의 차를 제곱해 주고 이를

각 Y에 해당하는 것마다 calculateMeanSquaredErrorPerInstance를 통해 따로 구해주고 calculateMeanSquaredError에서 이를 합하고 그 수만큼 나누어 평균을 구한다. 그리고 결과값은 매 numTraning 마다  이렇게 Train과 Test의 MSE가 각각 출력되며 마지막에 Graph와 List의 형태로 출력된다.

UML Diagram

|  |
| --- |
| **Analysis :: RegressionAnalysis** |
| +param:List = np.ones(numFeature,dtype=np.float32)  -verbose:Boolean = False |
| +train(Xs,Ys,iteration=10):float32  +searchParameters(Xs,Ys):float32  +test(Xs,Ys):float32  +calculateFirstDerivative(Xs,Ys):List  +calculateSecondDerivative(Xs,Ys):List  +calculateMeanSquaredError(Xs,Ys):float32  +calculateMeanSquaredErrorPerInstance(X,Y):float32 |

Attribute = param : parameter를 저장하는 list이다.

Verbose : verbose값이 True이면 매번 MeanSquaredError와 parameter의 값을 장황하게 출력하게 된다

Methods = train : parameter을 찾는 process인 searchParameters를 반복하고 calculateMeanSquaredError(Xs,Ys)를 통해 MSE값을 return한다.

searchParameters: calculateFirstDerivative, calculateSecondDerivative의 결과값을 가지고 parameter의 값을 갱신한다.

Test : MSE값을 return한다.

calculateFirstDerivative ,calculateSecondDerivative ,calculateMeanSquaredError ,calculateMeanSquaredErrorPerInstance : 각자 주어진 값을 calculate 한다.

