**기계학습 – 기법 간의 성능 테스트**

16010980 이우석

1. **성능 테스트에 사용된 기법**
2. 분류

* KNN classification
* Decision tree
* SVM classification

1. 회귀

* KNN regression
* Ridge regression
* LASSO

1. **성능 테스트 결과 표**
2. 분류

|  |  |  |  |
| --- | --- | --- | --- |
| 분류 문제 | | | |
|  | KNN classification | Decision tree | SVM classification |
| Accuracy score  (train\_data) | 0.975 | 1.0 | 0.983 |
| Accuracy score  (test\_data) | 0.967 | 0.9 | 0.933 |

1. 회귀

|  |  |  |  |
| --- | --- | --- | --- |
| 회귀 문제 | | | |
|  | KNN regression | Ridge regression | LASSO |
| MSE  (train\_data) | 0.739 | 0.529 | 1.339 |
| MSE  (test\_data) | 1.098 | 0.508 | 1.299 |

1. **의견**
2. 분류

* 분류에서는 KNN classification 기법이 3개의 분류 기법 중 가장 단순한 기법임에도 불구하고, test data 기준으로 가장 높은 점수를 받은 점이 의아했다. 이를 통해서 단순히 기법의 알고리즘이 복잡하다고 해서 성능이 좋은 것만은 아니라는 것을 알 수 있었다.

1. 회귀

* 회귀에서는 KNN regression 기법의 경우, train data 에 대한 MSE 와 test data 에 대한 MSE 가 꽤 많은 차이를 보인다. 아무래도 train data 에 대한 과대적합이 발생하지 않았나 싶다.
* Ridge regression 은 회귀 문제를 해결하는 데 있어 3개의 기법 중 가장 좋은 결과를 보여주었다. L2 규제화를 이용하는 Ridge regression 이 L1 규제화를 이용하는 LASSO 에 비해 굉장히 많은 성능 차이를 보였는데, 아무래도 L1 이 L2 보다 더 sparse 한 효과를 가진다고 하였으므로, 아마도 이 때문에 큰 성능 차이를 보인 것이 아닌가 생각한다.

1. **Google colab 노트북 링크 공유**
2. 분류

* KNN classification: <https://colab.research.google.com/drive/1gJvT9resAY9e3YLu474GXpezsR8UZhli?usp=sharing>
* Decision tree: <https://colab.research.google.com/drive/1Js9wgSozegviQZ03KMa6FiBPFvKgW-D3?usp=sharing>
* SVM classification: <https://colab.research.google.com/drive/1GuEkt7r6FE9wK0hLtnZ5s3k8dVCE7NSi?usp=sharing>

1. 회귀

* KNN regression: <https://colab.research.google.com/drive/1TYMSuJxgQLuxq6u8Acb0miW7ftL8JopW?usp=sharing>
* Ridge regression: <https://colab.research.google.com/drive/1OIIHPRqzq2dWnvotMwMTjqRnwiMG-_LU?usp=sharing>
* LASSO: <https://colab.research.google.com/drive/1BGDl_lb0iW55LZRH-r818CdZ21v29mPt?usp=sharing>