**Supervised learning model & accuracy test**

1. **Iris data Classification** 
   1. **Dataset :** IRIS dataset
   2. **Goal :** Classify species of iris(setosa, versicolor, virginica) with petal & sepal features
   3. **Result**

|  |  |
| --- | --- |
| **SVM linear model prediction** (1.0) | **SVM poly kernel model prediction** (.97) |
|  |  |
| **SVM rbf kernel model prediction** (.97) | **KNN model prediction** (.97) |
|  |  |
| **Decision Tree prediction** (.95) | **Random Forest prediction** (.95) |
|  |  |

1. **MNIST data Classification** 
   1. **Dataset :** MNIST (use only 10% of training dataset)
   2. **Goal :** Classify numbers(0~9) with picture file of hand-written number(pixel data)
   3. **Result**

|  |  |
| --- | --- |
| **SVM linear model prediction** (.91) | **SVM poly kernel model prediction** (.95) |
|  |  |
| **SVM rbf kernel model prediction** (.91) | **KNN model prediction** (.94) |
|  |  |
| **Decision Tree prediction** (.78) | **Random Forest prediction** (.94) |
|  |  |

1. **Conclusion**
   1. **Iris :** 간단한 데이터셋의 경우 linear SVM이 가장 정확한 prediction을 나타낼 수도 있음.

**accuracy 높은 순서 : linear > poly, RBF, KNN > RF, tree**

* 1. **MNIST :** SVM의 poly kernel의 경우 2차원의 점을 3차원의 점으로 변환하므로, 기존 pixel값(z축으로 활용)이 0~256이었을 경우 높은 prediction을 보이나, 해당 값을 256으로 나누면(0~1) 높은 prediction을 보여주지 못하는 모습을 보였음. 반면에 rbf커널은 그와 상반된 모습을 보인다.  
     **accuracy 높은 순서 : poly > RF > KNN > RBF > linear > tree**
  2. **종합 :** linear의 경우 label 및 feature의 개수가 많아질수록 낮은 정확도를, tree는 꾸준히 높은 정확도를 보임