ARKLAB intern training: 2023.01.05

Machine learning

YuBin, Hong (System team)

School of Life Science, HGU

Contents

1. 머신러닝이란?

2. 왜 머신러닝을 사용하는가? 3. 적용 사례

4. 머신러닝 시스템의 종류 5. 머신러닝의 주요 도전 과제 6. 테스트와 검증

To do list

모든 데이터 과학자가 꼭 알아야할 여러 가지 기초 개념과 용어 설명

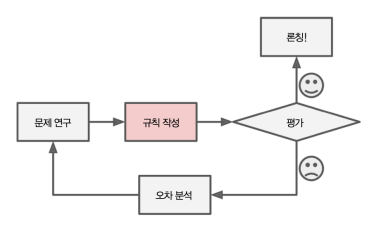
**1.1 ?**

(spam) (ham)

(trainingset):() (traininginstance):

**1.5**

**1.2 ?**

****

.

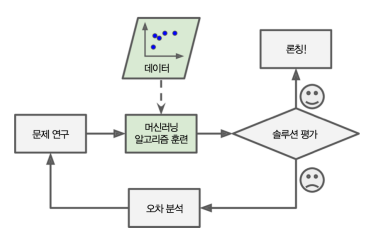
:

:

:

,

**:**

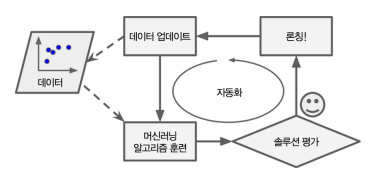
****,.

:

:,

**:**

'ForU'

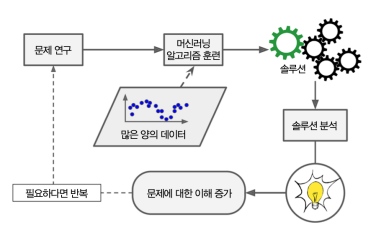


**MLOps(Machine Learning**

**Operations, )** .

: MLOps:

(https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery and-automation-pipelines-in-machine-learning)

( **data mining**)

**1.3**

:

:

( ): :

:

: (chatbot)

:

:

:

:

:

:

:(bot)

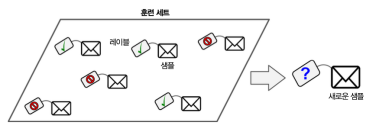
**1.4**

.

+ +

**(label)**

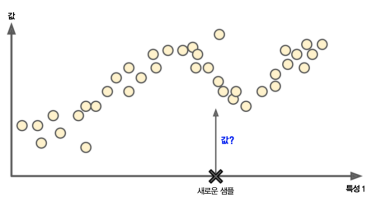
**(target)**.



:

:,

():



(target)

:

:,,

:

k- (k-NNs)

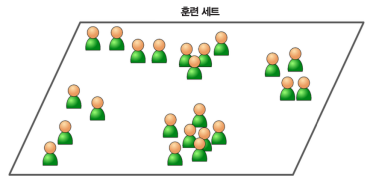
(linear regression)

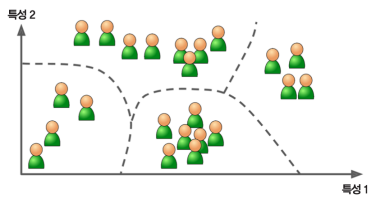
(logistic regression)

(support vector machines, SVCs)

(decision trees) (random forests) (neural networks)

: , SVM



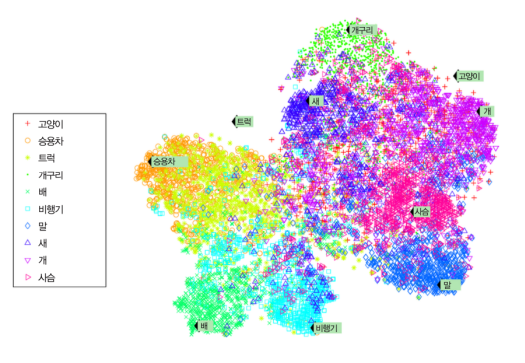
: , , , , , SF,

\* k-

\* DBSCAN \*

2D 3D

2



:

.''

.

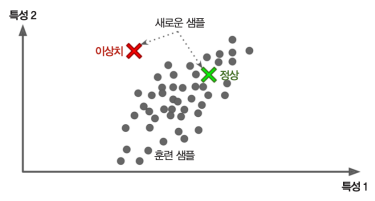
:

\* (PCA) \* PCA

\*

\* t-SNE

**(anomaly detection)**

****.

(outliers)

**(novelty detection)** ' '(clean)

**vs.**

:1%

.

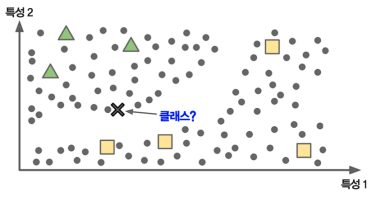
.

- SVM Isolation Forest

:

.

Apriori() Eclat()

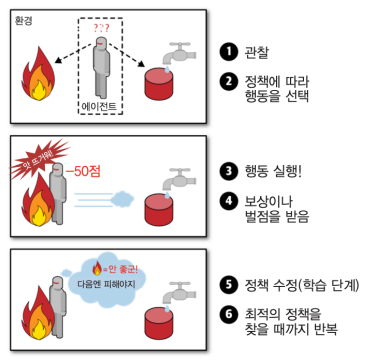


:.. :X.

(deepbeliefnetwork,DBN)

(restrictedBoltzmannmahcine), .

( )



(DeepMind) (AlphaGo)

**vs.**

**(batch learning)**

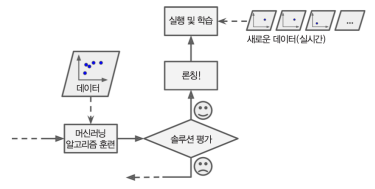
(cpu, gpu, , )

. MLOps

(https://cloud.google.com/solutions/machine-learning/mlops-continuous delivery-and-automation-pipelines-in-machine-learning)

**(onlinelearing)**

(,mini-batch)



.

:



**vs.**

**(generalization)**

:''

k-NN

:X

.



: X

: OECD 1 GDP(1 )



1GDP. ()

:

=1GDP

:

:

:()



:

:





:

:



:

:

:





**1.6**

:

(80%) (20%)

: .

:

.

:

:

**(hyper-parameter)**

( holdout) ( ):

:

:

:

-:,

,-

:

**(Nofreelunch)**

.

.