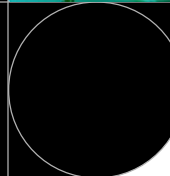


Chapter 12

인공신경망 이용 주가예측

Sejong Oh

Bio Information technology Lab.



Content

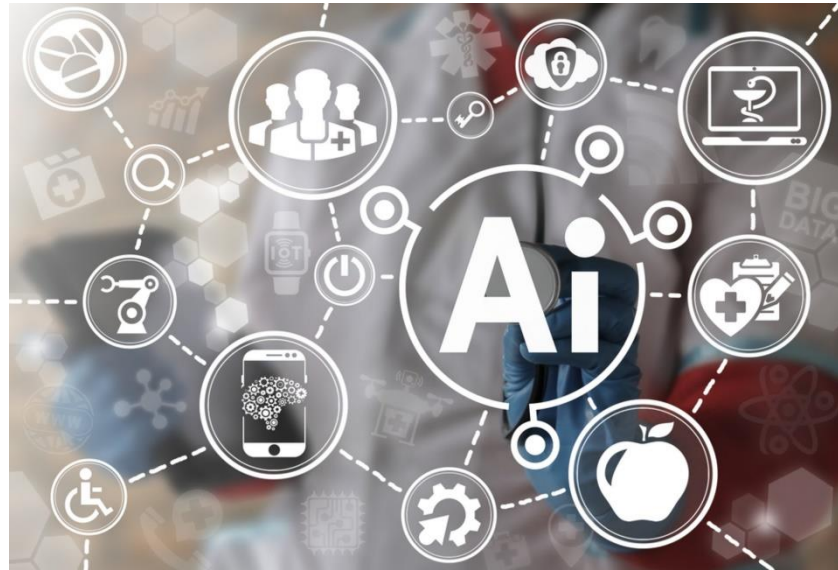
- 인공지능 개요
- 인공 신경망 모형
- 인공신경망을 이용한 주가 예측

이 slide 는

R로 배우는 코딩, 장용식/강희구, 생능출판사
를 참조하였음

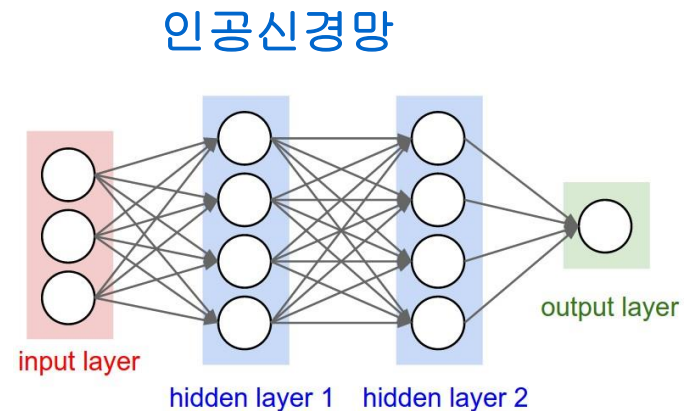
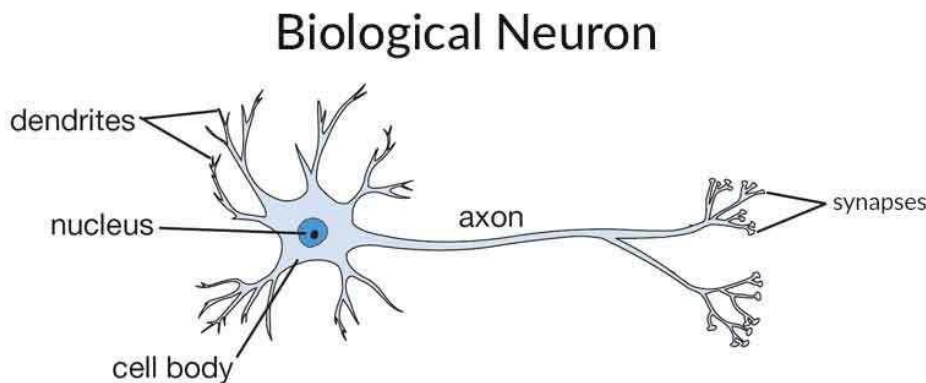
1. 인공지능 개요

- 인공지능: 기계에 의해 나타나는 지능
 - 지식 체계를 규칙(if then else)으로 표현하는 전문가 시스템, 자연언어에서의 애매모호함을 정량적으로 표현하는 퍼지 이론, 컴퓨터에 인공적인 학습이 가능한 지능을 부여하는 기계 학습, 자연의 진화 과정을 통해 특정한 문제의 적절한 답을 찾는 유전 알고리즘 등의 기술로 가능

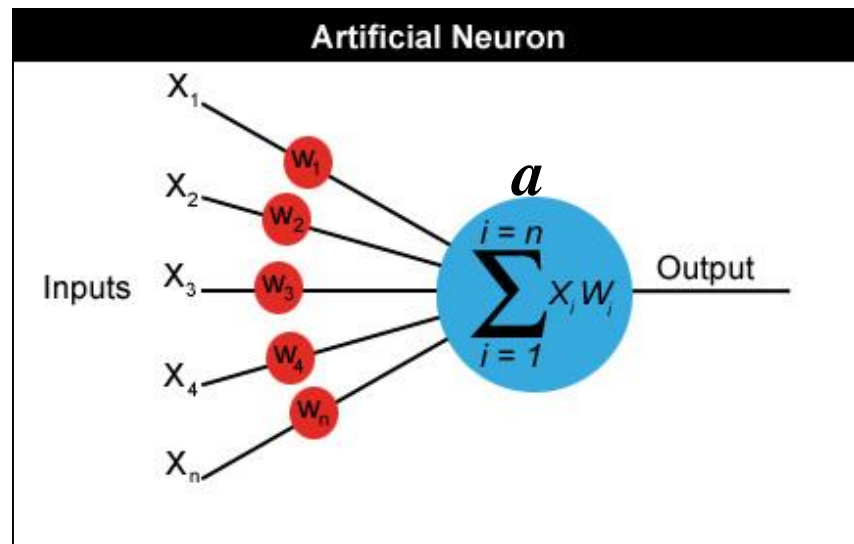


1. 인공지능 개요

- 인공 신경망: 기계 학습 분야에서 연구되는 학습 알고리즘의 하나
 - 시계열 자료의 예측, 분류, 패턴 인식, 제어 분야 등 다양하게 응용
 - 대부분 통계적으로 접근하여 확률을 계산하는 다른 기계 학습 기술과 달리, 인간의 뇌의 구조를 모방
 - 인공 신경망은 수상돌기를 입력으로, 신경세포체를 노드node로, 축색을 출력으로 흉내 내고, 시냅스는 노드 간 가중값을 갖는 연결로 네트워크 구성.
 - 입력 데이터에 대해 목표치를 출력하려는 학습 과정을 거치는 동안 각 연결망의 가중치가 달라짐
 - 구글 딥마인드의 알파고



2. 인공 신경망 모형



$$a = x_1 w_1 + x_2 w_2 + x_3 w_3 \dots + x_n w_n$$

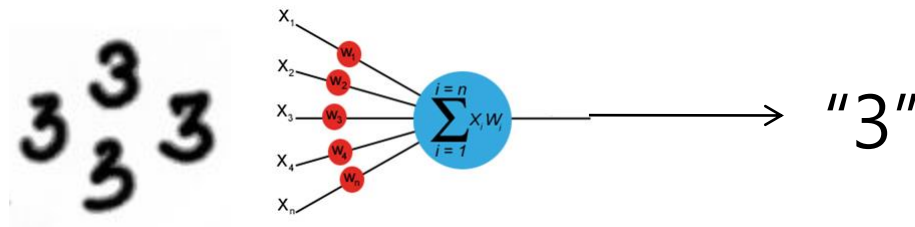
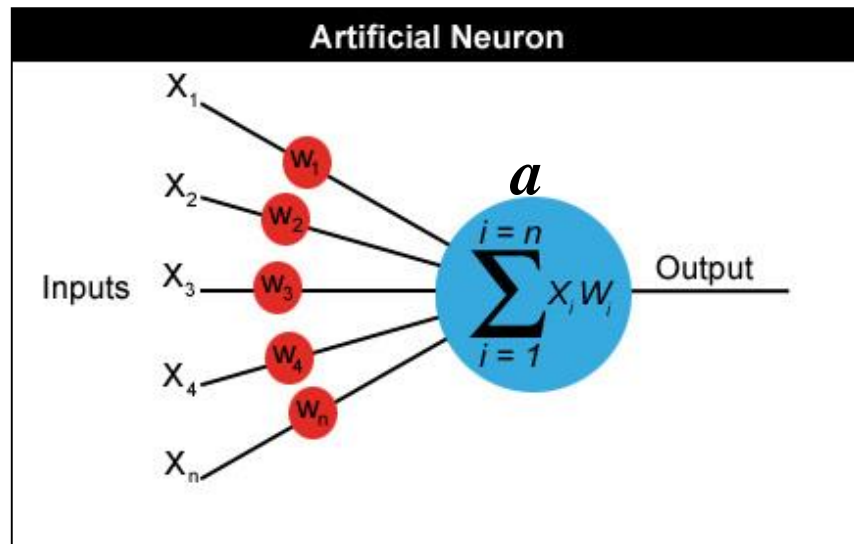
$$a = \sum_{i=0}^n w_i x_i$$

$$\text{output} = \begin{cases} 0 & \text{if } a < \text{threshold} \\ 1 & \text{if } a \geq \text{threshold} \end{cases}$$

2. 인공 신경망 모형

- Idea

- If we adjust W_i value, we can get output which we want



3. 인공 신경망을 이용한 분류

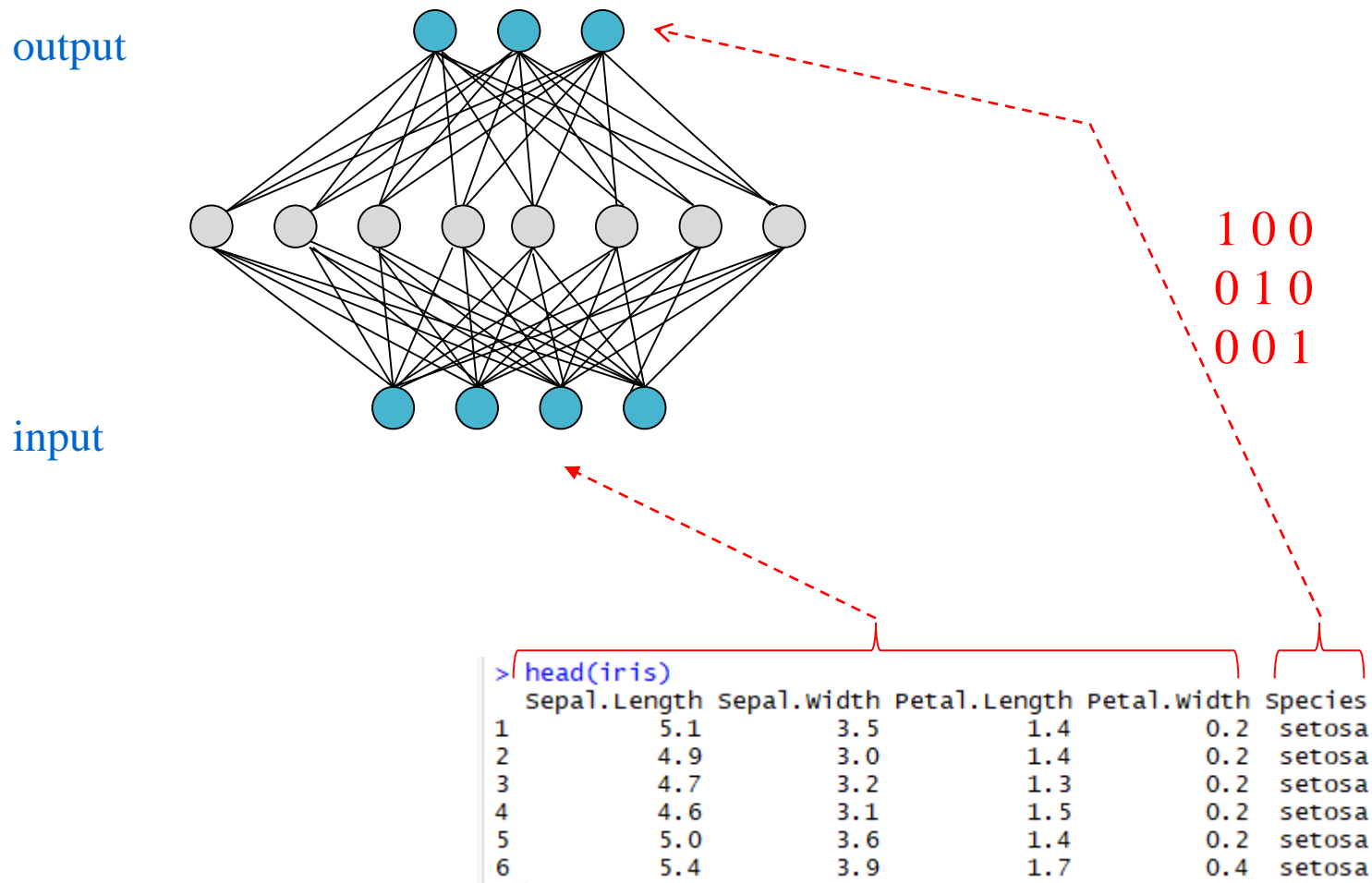
- nnet 함수 (in "nnet" package)
- Training

```
model = nnet(tr, target, size, rang, decay, maxit)
```

- tr : training data (matrix or vector)
- target : output of NN (class 정보)
- size : hidden node 수
- rang : Initial random weights on [-rang, rang].
- decay : parameter for weight decay (default: 0)
- maxit : maximum number of iterations (default: 100)
- Test

```
pred = predict(model, ts)    # ts : test data
```

- Neural net for iris



Practice

```
require(nnet)
data(iris)

train = iris[,-5]
test = train
targets = class.ind(iris[,5])

model = nnet(train, targets, size = 8,
             rang = 0.1, decay = 5e-4, maxit = 200)

# test with train data
pred <- predict(model, train)
```

```
> pred
```

	setosa	versicolor	virginica
[1,]	0.989901184	0.0099054336	9.622787e-05
[2,]	0.989423540	0.0105653736	9.621525e-05
[3,]	0.989727770	0.0101436494	9.622322e-05
[4,]	0.989156013	0.0109401681	9.620843e-05
[5,]	0.989928503	0.0098680504	9.622860e-05
[6,]	0.989830220	0.0100027224	9.622595e-05
[7,]	0.989659249	0.0102382112	9.622140e-05

Practice

```
# calculate accuracy
predClass = max.col(pred)
testClass = max.col(targets)
acc = mean(predClass == testClass)

# check accuracy
acc
table(predClass, testClass)
```

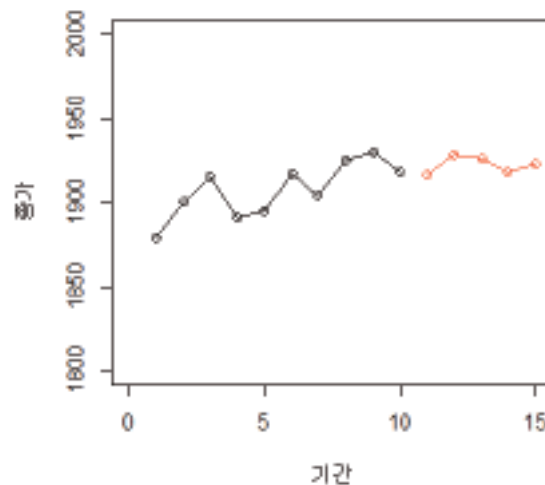
4. 인공 신경망을 이용한 주가 예측

한국 거래소에서 6개월간(2016. 1. 1~2016. 6. 30)의 일별 주가 데이터를 다운로드하고 인공 신경망 모형을 만들어 학습 과정을 거친 후, 향후 5일 동안의 주가를 예측하는 방법을 살펴보자.

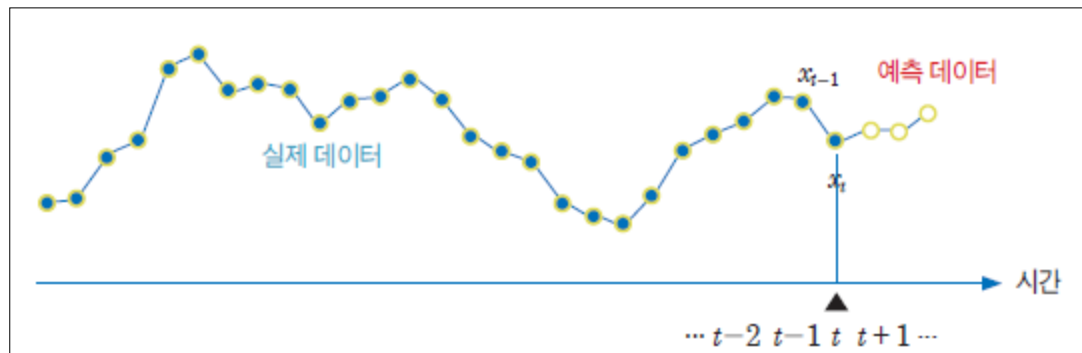
☞ 인공 신경망 모형

입력 노드 수	10 개
은닉층 수	1 개
은닉층 노드 수	20 개
출력 노드 수	5 개

미리 보기



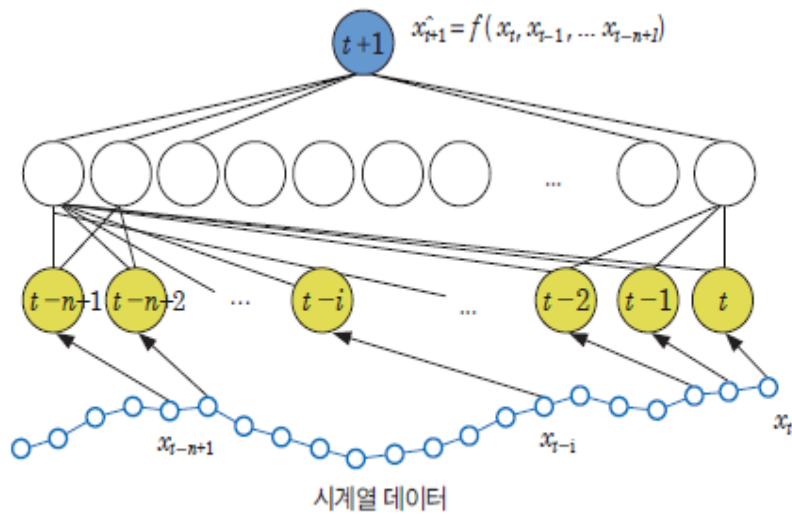
4. 인공 신경망을 이용한 주가 예측



출력층

은닉층

입력층

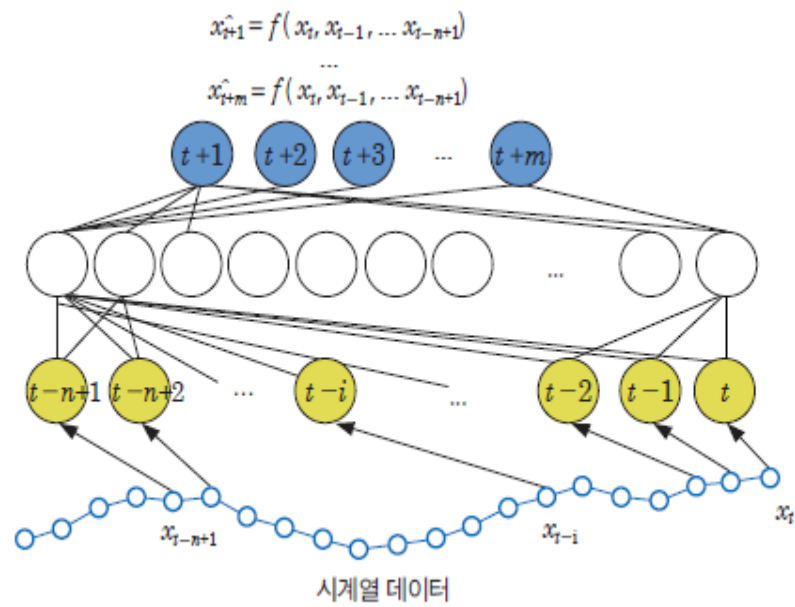


1단계 예측 모형

출력층

은닉층

입력층



다단계 예측 모형

시계열 데이터 준비

- 한국 거래소 www.krx.co.kr의 통계 정보로 제공되는 코스피(KOSPI) 주식 지수의 증가 시계열 데이터 다운로드

www.krx.co.kr

The screenshot shows the KRX website's main navigation bar with the following tabs: KRX 시장, 시장정보, 상장공시, 시장감시, 규정/제도, 금융교육, and KRX 소개. The '시장정보' (Market Information) dropdown menu is open, displaying a list of market data categories: 시장동향, 증권시장지, 지수, 주식, 채권, 파생, 일반상품, ETF/ETN/ELW, 해외연계시장, and 통계. The '통계' (Statistics) item is highlighted with a red rectangular box. Below the menu, a banner features a line graph and the text '세상의 모든 가치가 만나 더 큰 가치를 이룰 수 있도록 한국거래소가 끊임없이 혁신해 나가겠습니다.' (The world's all values meet to create even greater value so that the Korea Exchange can continuously innovate and move forward.)

80001 일자별지수



계열구분	<input type="radio"/> 전체 <input type="radio"/> KRX <input checked="" type="radio"/> KOSPI <input type="radio"/> KOSDAQ <input type="radio"/> 파생기타 <input type="radio"/> 상품
업종구분	코스피
조회기간	20170101 ~ 20170630 <input type="button" value="달력"/> <input type="button" value="1일"/> <input type="button" value="1개월"/> <input type="button" value="6개월"/> <input type="button" value="1년"/>

Q 조회

차트 | Excel | CSV

1

2

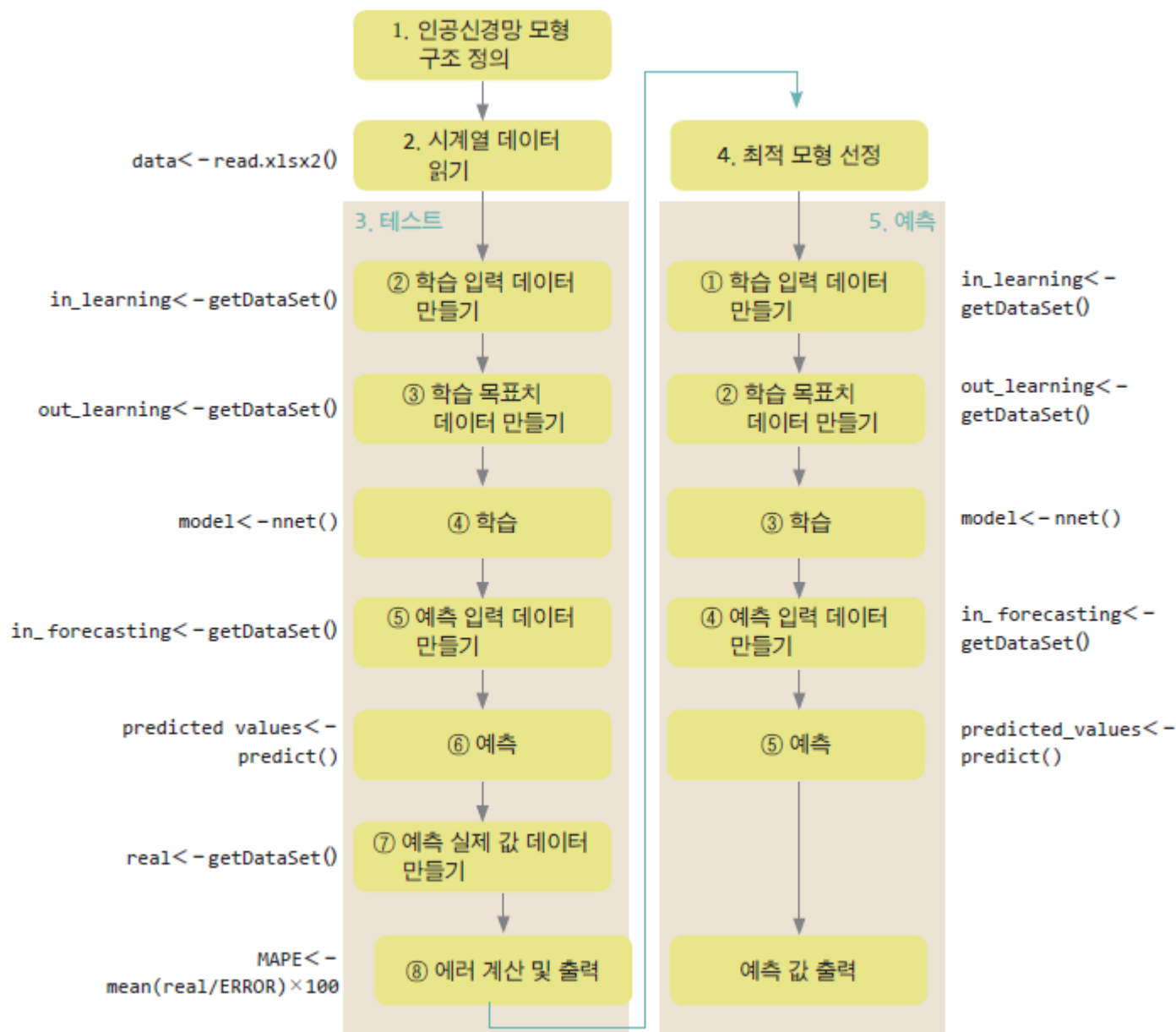
년/월/일	현재지수	대비	등락률(%)	시가지수	고가지수	저가지수
2017/06/30	2,391.79	▼ 3.87	-0.16	2,382.15	2,391.81	2,381.15
2017/06/29	2,395.66	▲ 13.10	0.55	2,396.81	2,402.80	2,385.81
2017/06/28	2,382.56	▼ 9.39	-0.39	2,382.91	2,390.89	2,373.56
2017/06/27	2,391.95	▲ 3.29	0.14	2,386.76	2,397.14	2,385.47
2017/06/26	2,388.66	▲ 10.06	0.42	2,381.09	2,390.70	2,370.66
2017/06/23	2,378.60	▲ 8.23	0.35	2,371.54	2,380.94	2,360.37
2017/06/22	2,370.37	▲ 12.84	0.54	2,364.22	2,370.39	2,351.53
2017/06/21	2,357.53	▼ 11.70	-0.49	2,358.23	2,359.97	2,345.83
2017/06/20	2,369.23	▼ 1.67	-0.07	2,382.54	2,382.54	2,360.86
2017/06/19	2,370.90	▲ 9.07	0.38	2,364.02	2,377.20	2,350.83

- kospi.csv

	A	B	C	D	E	F	G	H	I
1	년/월/일	현재지수	대비	등락률(%)	시가지수	고가지수	저가지수	거래량(천주)	거래량(천주)
2	2017-06-30	2,391.79	-3.87	-0.16	2,382.15	2,391.81	2,381.40	343,051	343,530
3	2017-06-29	2,395.66	13.1	0.55	2,396.81	2,402.80	2,393.57	445,044	445,132
4	2017-06-28	2,382.56	-9.39	-0.39	2,382.91	2,390.89	2,380.75	361,053	361,177
5	2017-06-27	2,391.95	3.29	0.14	2,386.76	2,397.14	2,383.47	357,934	358,034
6	2017-06-26	2,388.66	10.06	0.42	2,381.09	2,390.70	2,377.88	325,537	325,674
7	2017-06-23	2,378.60	8.23	0.35	2,371.54	2,380.94	2,369.17	256,067	256,259
8	2017-06-22	2,370.37	12.84	0.54	2,364.22	2,370.39	2,354.10	299,093	299,232
9	2017-06-21	2,357.53	-11.7	-0.49	2,358.23	2,359.97	2,346.19	360,030	360,345
10	2017-06-20	2,369.23	-1.67	-0.07	2,382.54	2,382.54	2,365.45	360,475	360,734
11	2017-06-19	2,370.90	9.07	0.38	2,364.02	2,377.20	2,356.84	290,383	290,945
12	2017-06-16	2,361.83	0.18	0.01	2,364.20	2,365.37	2,355.29	328,160	329,179
13	2017-06-15	2,361.65	-10.99	-0.46	2,373.36	2,378.04	2,350.37	334,199	334,886
14	2017-06-14	2,372.64	-2.06	-0.09	2,384.54	2,387.29	2,366.92	392,027	392,998
15	2017-06-13	2,374.70	16.83	0.71	2,358.92	2,375.81	2,358.92	321,589	322,082
16	2017-06-12	2,357.87	-23.82	-1	2,370.69	2,374.70	2,353.35	326,087	327,234
17	2017-06-09	2,381.69	18.12	0.77	2,371.22	2,385.15	2,367.81	519,732	520,558
18	2017-06-08	2,363.57	3.43	0.15	2,361.64	2,365.52	2,347.62	416,713	417,080
19	2017-06-07	2,360.14	-8.48	-0.36	2,364.96	2,369.43	2,356.71	375,909	376,208
20	2017-06-05	2,368.62	2.1	0.12	2,376.66	2,376.92	2,362.61	276,260	276,412

이 데이터를 이용함

학습과 예측 과정



1단계: 인공 신경망 구조 정의

인공 신경망 구조

```
INPUT_NODES <- 10  
HIDDEN_NODES <- INPUT_NODES * 2  
OUTPUT_NODES <- 5  
ITERATION <- 500
```

1. 인공 신경망 모형의 구조 정의

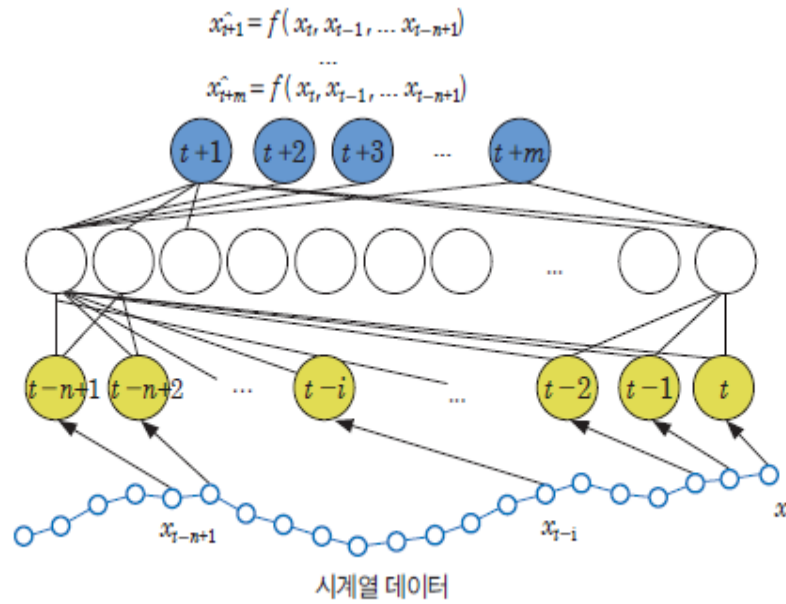
- 입력 노드 수
- 은닉층의 노드 수
- 출력 노드 수
- 학습 반복 수

미래주가지수

출력층

과거주가지수

입력층



다단계 예측 모형

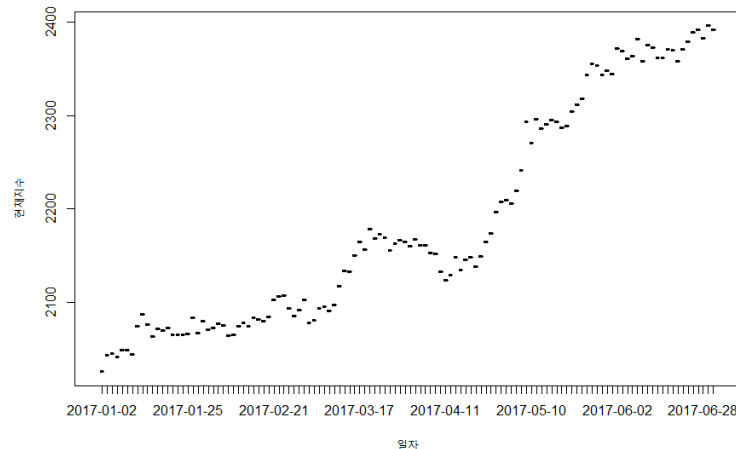
2단계: 시계열 데이터 읽기

```
library(nnet)

setwd("c:/work")
data <- read.csv("kospi.csv")

names(data)[1:2] = c("ymd", "current")
data$current <- gsub(",", "", data$current) #콤마제거
df <- data.frame(ymd=data$ymd,
                 current=as.numeric(data$current))
plot(df$ymd, df$current, xlab="일자", ylab="현재지수",
     type="o")
```

```
> head(df)
      ymd current
1 2017-06-30 2391.79
2 2017-06-29 2395.66
3 2017-06-28 2382.56
4 2017-06-27 2391.95
5 2017-06-26 2388.66
6 2017-06-23 2378.60
```



3단계: 테스트

- Training/test dataset 생성
- 인공신경망 학습
- 테스트 데이터로 주가예측
- 에러계산 : 실제 주가지수와 예측된 주가지수의 차이

3단계: 테스트 (1)

- 학습/테스트 데이터 생성

```
df <- df[order(df$ymd),] # sort date

# create traing data
train_learning <- NULL
train_result <- NULL
for (i in 1:108) {
  train_learning <- rbind(train_learning,
                          df$current[i:(i+9)])
  train_result <- rbind(train_result,
                        df$current[(i+10):(i+14)])
}
```

```
> dim(train_learning)
[1] 108 10
> dim(train_result)
[1] 108 5
```

- 학습/테스트 데이터 생성

```
# create test data
test_learning <- NULL
test_result <- NULL
for (i in 51:100) {
  test_learning <- rbind(test_learning,
                        df$current[i:(i+9)])
  test_result <- rbind(test_result,
                      df$current[(i+10):(i+14)])
}
```

```
> dim(test_learning)
[1] 50 10
> dim(test_result)
[1] 50 5
```

3단계: 테스트 (2)

- 학습 실시

```
# training & create a model
```

```
HIDDEN.NODES = 20
```

```
ITERATION = 500
```

```
model <- nnet(train_learning, train_result,  
              size=HIDDEN.NODES,  
              linout=TRUE, rang=0.1, decay = 5e-4,  
              skip=TRUE, maxit=ITERATION)
```

```
> model <- nnet(train_learning, train_result,  
+               size=HIDDEN.NODES,  
+               linout=TRUE, rang=0.1,  
+               skip=TRUE, maxit=ITERATION)  
# weights: 375  
initial value 2630438619.060554  
iter 10 value 540729101.448966  
iter 20 value 14478322.234670  
iter 30 value 1314021.972509  
iter 40 value 195314.379409  
iter 50 value 193362.595098  
final value 193341.205245  
converged  
>
```

- liout : 신경망의 output 이 linear
- rang : weight matrix 초기화 $[-rang, rang]$
- skip : skip-layer connection

3단계: 테스트 (3)

- 주가예측 테스트 실시

```
# test using model
predicted <- predict(model, test_learning,
                      type="raw")
predicted
```

```
> head(predicted)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 2167.182 2168.476 2170.867 2172.615 2175.805
[2,] 2166.436 2169.404 2171.038 2173.961 2177.821
[3,] 2163.281 2164.133 2166.833 2169.943 2172.807
[4,] 2168.819 2172.472 2175.889 2179.419 2182.387
[5,] 2165.120 2168.088 2170.701 2173.210 2176.887
[6,] 2164.263 2167.515 2169.153 2172.927 2175.352
```

3단계: 테스트 (4)

- 예측 에러 계산

```
err <- abs(test_result - predicted)
MAPE <- mean(err/test_result)*100

mean(err)
MAPE
```

```
> mean(err)
[1] 17.8212
> MAPE
[1] 0.7897159
~
```

← 실제주가지수와 예측 주가지수의 차이 평균

← 실제주가지수 에서 에러가 차지하는 비율

MAPE: Mean Absolute Percentage Error

별로 좋은 모델은 아님

4단계: 최적모형 선정

- 인공 신경망의 입력 노드, 출력 노드, 은닉층의 노드 수에 따라 다양한 모형 구조가 가능하고, 같은 구조라 하더라도 반복 학습 수에 따라 서로 다른 모형이 됨
- 따라서, 여러 모형들을 비교하여 MAPE가 가장 작은 모형을 선택해서 원본 전체 1~127번 데이터에 대해 다시 학습하고 최종적으로 미래 기간을 예측함

Practice 1

- 미국 달러 환율데이터를 가지고 미래 환율을 예측해 보시오
 - 학습기간 : 2017.1.1 ~2017.9.30
 - 예측기간 : 2017.10.1~10.5
 - 환율데이터 다운로드
 - 하나은행(<https://www.kebhana.com>)-> 외환포탈 -> 고시환율 -> 환율변동

환율변동

> 2014.3.4 이후 송국 위안화 환율은 홍콩시장에서 거래되는 위안화 환율(CNH)을 고시합니다.

조회구분	<input type="radio"/> 일중환율변동 <input type="radio"/> 월중환율변동 <input checked="" type="radio"/> 기간환율변동
조회기간	2017-01-01 ~ 2017-09-30
통화	USD:달러(미국) ▼
고시회차	최초 ▼

금융계산기

조회

텍스트로저장 엑셀로저장 인쇄

- 엑셀로 예측치와 실제 값이 얼마나 차이나는지 비교해 보시오