

인공지능 — 머신러닝



거시경제를 예측하다

SVM 환율예측 모델링

정원일 Economist

02 3770 3526

wonil.jeong@yuantakorea.com

김호정 Research Assistant

02 3770 3630

hojung.kim@yuantakorea.com

인공지능 | 머신러닝 거시경제를 예측하다

SVM 회귀분석, 머신러닝



정원일

Economist

02 3770 3526
wonil.jeong@yuantakorea.com



김호정

Research Assistant

02 3770 3630
hojung.kim@yuantakorea.com

소개글

- 기술의 발전에 따라 인공지능을 통한 머신러닝, 그리고 이를 적용한 분야는 기하급수적으로 늘어나고 있습니다. 이러한 기조는 금융시장에서도 예외가 될 수 없는 상황입니다. 이와 같은 상황에 대한 대응 차원에서 본 연구에서는 머신러닝 기법을 환율 예측에 적용하였습니다.
- 이는 전통적인 예측방법인 경제학이론+통계학기법을 통한 예측과는 별개로 해석할 필요가 있습니다. 논리로 설명하기 어려우나 각종 감정과 고정관념을 배제하였기 때문에 하나의 객관적 현상을 토대로 학습이 되었기 때문입니다.
- 다만 아직까지는 그 정확성 측면에서 여러 연구논문에서 지적되었듯이 편차가 상대적으로 높은 수준입니다. 따라서 변수사용 및 모델링 측면에서 정합성을 높이기 위한 연구가 지속되고 있고, 본 연구 역시 학습능력을 업그레이드해야 할 필요가 있습니다.
- 본 연구를 통하여 금융시장에서의 머신러닝 기법 도입 가능성을 확인할 수 있었습니다. 지속적인 모델링과 추정기법 개선이 수행될 것이며, 추정된 환율예측 결과는 필자의 주간 경제노트 자료에 매주 업데이트하여 결과를 제시할 예정입니다.

Summary

1. SVM의 개념

- 서포트벡터머신(SVM)은 머신러닝의 분야 중 하나로 주로 분류를 위해 사용. SVM을 활용한 환율 예측 방법은 환율에 영향을 줄 수 있는 요소들을 특성으로 선정하여 패턴들을 학습하는 방식. 회귀분석과 비교할 때, 회귀분석의 경우에는 모형 설정상 자유도의 문제 발생. 그러나 SVM은 오히려 많은 투입변수를 활용할 때 좋은 성과.

2. SVM 기반 환율예측 모델링

- 본 연구에서 input data로 활용하는 변수의 category는 금리, 주가, 주가변동, 원자재, 환율로 구성됨. 금리로 활용되는 지표는 각국의 장기금리(10년물), 단기금리(2년물), 물가채금리 등이 있음. 원자재가격은 유가와 금 가격으로 구성됨. 주가는 주요국의 대표 지수를 사용하였으며, 변동성지수가 동시에 포함됨. 또한 환율은 달러대비 환율로 구성됨.
- SVM은 주어진 정보를 활용하여 지도학습법을 진행하고, 이를 토대로 back-testing을 진행하는 방법. 따라서 전체 샘플을 training set과 test set으로 구분하는 작업이 필요함. 본 연구에서는 전체 225개의 샘플 중 4/5인 180개 샘플을 training set으로 선정하고, 나머지 45개의 샘플에 대해서 백테스트를 수행하였음.

Summary

3. 머신러닝을 통한 환율예측 결과

- 추정의 정확성을 판단하기 위하여 iteration 회차별 추정값과 실제값의 일치성을 판단. 1000회의 반복시행을 통하여 추정된 전체 test data에 대한 정확성을 판단한 결과는 정확한 추정을 하는 경우가 65.6% 수준인 것으로 나타남. 실제 t+1기의 환율 방향성에 대한 분포의 평균값으로 다음주의 환율이 상승할 가능성을 22.1%로 정의하고 하락할 가능성을 77.9%로 제시할 수 있음.
- 백테스트를 위하여 2018년 1월 이후 현재까지의 환율수준을 상승과 하락 국면으로 구분하여 실제값과 추정값의 차이를 분석함. 63회의 추정 중 40회의 추정이 정확히 이루어졌기 때문에 약 63.5%의 정확도를 가지는 것으로 해석됨. 백테스트에 대한 함의로, 환율이 안정적으로 횡보하는 국면은 다음주의 환율 방향성을 예측할 때 정확도가 높은 것이 확인된 반면 structural break 등이 발생한 2018년 6월~7월, 2018년 10월~11월 등의 국면에서는 불일치하는 모습.

4. 함의와 응용, 그리고 보완과제

- 환율예측에 머신러닝을 적용하였다는 점에서 최신 트렌드를 접목. 반복시행을 통한 정확성이 확보되었으며, 65.6%의 hit ratio 달성. 추정값의 일치국면과 불일치국면의 특징적 현상을 도출할 수 있었음. 금융지표의 국면전환, 가격변수와 거시경제변수의 추세 및 방향성을 판별하는 모델링으로 확장이 가능한 반면, 원인과 결과 사이의 black box인 점은 한계점으로 작용할 수 있음.



Contents

Part. 1	SVM의 개념	7
Part. 2	SVM 기반 환율예측 모델링	13
Part. 3	머신러닝을 통한 환율예측 결과	21
Part. 4	함의와 응용, 그리고 보완과제	35
Appendix	머신러닝의 소개	39



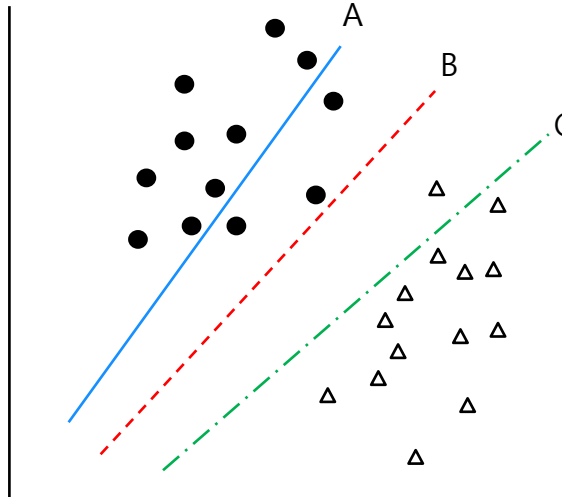
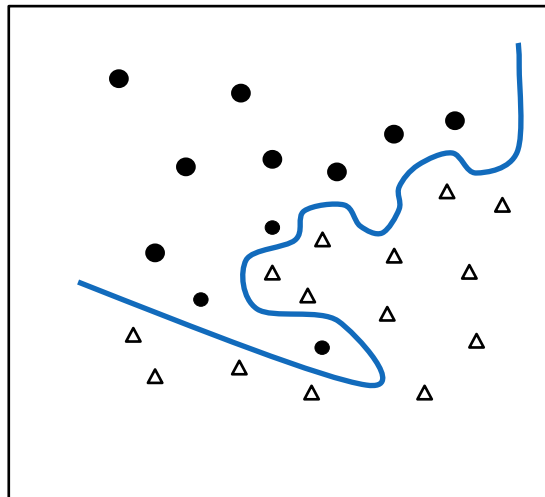
Part. 1

SVM의 개념

서포트 벡터 머신(support vector machines)이란?

- 서포트벡터머신(SVM)은 머신러닝의 분야 중 하나로 주로 분류를 위해 사용되며 서로 다른 두 집단에 속한 데이터들에 관한 정보를 얻었을 때, SVM 알고리즘은 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 집단에 속할지를 판단하는 알고리즘.
- SVM 알고리즘은 두 집단의 데이터를 가상의 벡터 공간에 사상(Mapping)시키고, 분류 기준을 공간상에 표현하고, 이후 새로운 데이터가 다시 공간에 진입했을 때, SVM 모델은 새로운 데이터의 위치에 따라 그 데이터가 어느 집단에 속하는지 판단.

일반기계학습(좌)의 일반화 처리와 SVM(우)의 일반화 처리

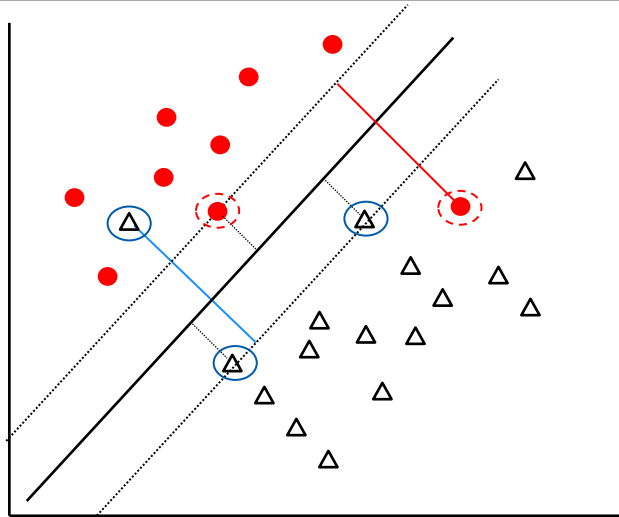


- 학습집합 구간에서는 B와 C는 무차별함
- 테스트집합(일반화 능력 측면) 여백이 더 큰 C가 B보다 우수함
- SVM은 일반화 능력에 있어 뛰어난 결과를 보이는데, 일반화 능력은 학습이 완료된 후 알고리즘의 성능을 평가하는 척도로 활용

선형 서포트 벡터 머신(linear support vector machines)

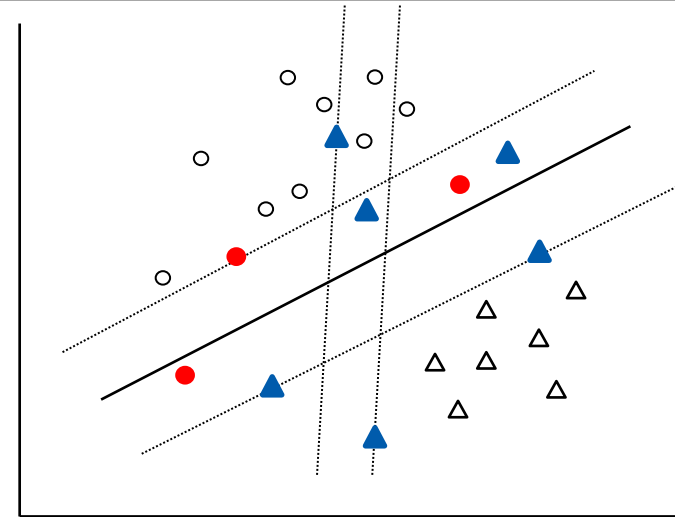
- SVM의 기본 모델은 선형형태를 기본으로, 오분류(Miss-Classification)없이 선형식(Linear Equation)으로 데이터가 완전히 분리되는 경우를 상정함. 이때 각 데이터 집단의 경계에 있는 점들을 서포트 벡터(Support Vectors)라고 하며, 서로 다른 집단의 서포트 벡터 사이의 직선거리를 마진(Margin)이라고 함.
- SVM 모델에서는 마진이 최대화되도록 선형식이 결정되고 이를 최대 마진 초평면(Maximum-Margin Hyperplane)이라고 함. 최대 마진 초평면은 새로운 데이터를 분류해주는 기준이 됨.
- 오분류 없이 데이터를 완전히 선형 분리할 수 없는 경우에는 오분류를 인정하되 오분류 정도를 최소화하는 방향으로 선형 SVM이 결정됨. 이때 SVM 모델은 '최대 마진', '오분류 최소화' 두 가지의 목적을 적절히 고려하며 선형식을 결정함.

선형 SVM(오분류가 없는 경우)



자료: 유안타증권 리서치센터

선형 분리가 불가능한 경우의 SVM

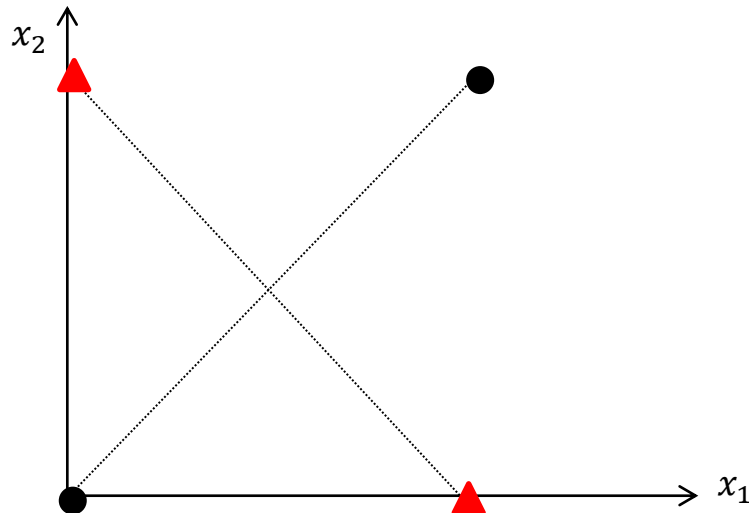


자료: 유안타증권 리서치센터

비선형 서포트 벡터 머신(Non-linear support vector machines)

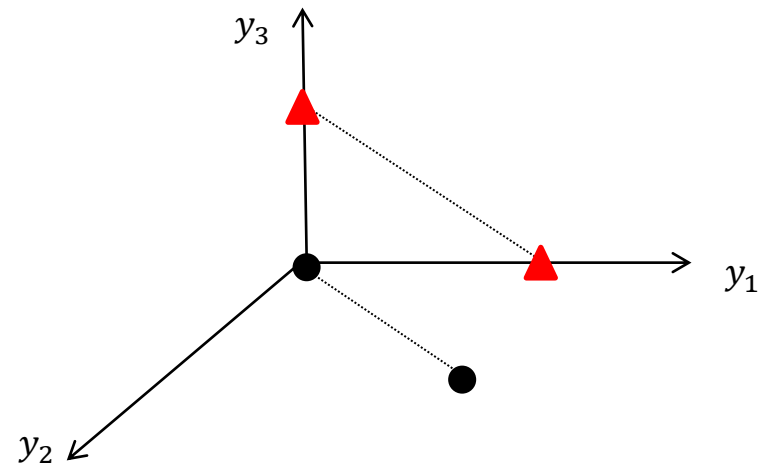
- Boser et al.(1992)에서는 최대 마진 초평면을 구하는 문제에 커널 트릭(Kernel Trick)을 적용하는 비선형 분류가 제안되었음. 이를 비선형 SVM이라고 하는데 '고차원 공간으로의 데이터 이동'이 핵심.
- 고차원 공간으로의 데이터 이동을 통한 비선형 SVM이 분류 성과를 더 높일 수 있음. 하지만 실제로는 고차원 공간으로의 사상을 가능하게 해주는 함수를 매번 찾을 수 없음.
- 하지만 고차원 공간으로의 이동을 위해 커널 함수(Kernel Function)를 이용. 커널 함수를 통해 비선형 SVM을 매핑하여 선형 분리를 가능하도록 함.

비선형하에서 원래의 공간 $x=(x_1, x_2)$



자료: 유안타증권 리서치센터

고차원으로 매핑된 공간 $\phi(x)=(y_1, y_2, y_3)=(x_1^2, \sqrt{2}x_1, x_2^2)$

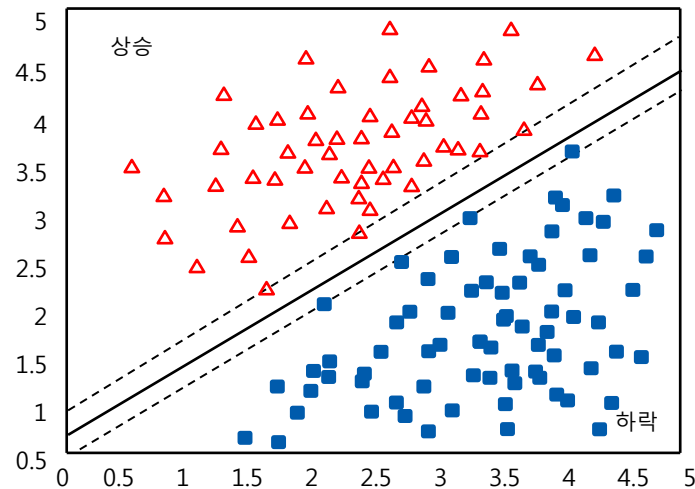


자료: 유안타증권 리서치센터

서포트 벡터 머신(support vector machines)을 통한 환율 예측

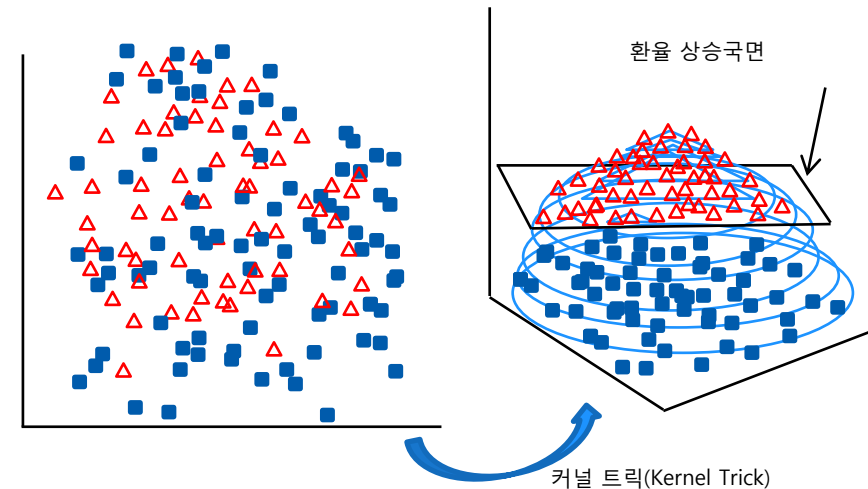
- SVM을 활용한 환율 예측 방법으로, 환율에 영향을 줄 수 있는 요소들을 특성으로 선정하여 패턴들을 학습하는 것임. 가장 쉽게 일일단위로 수집할 수 있는 채권금리, 주가, 원자재 등을 고려하고 환율이 상승할지 하락할지를 예측한다고 하면, 학습할 데이터를 초평면에 늘어놓으면 2차원에서 표현할 수 있고, 만약 2차원으로 표현이 불가능하다면 3차원으로 표현할 수 있음.
- 그 중에 증가하는 것에 1, 아니면(0)로 마크를 하고 학습함. 마크는 지도학습에서 결과값을 주는 것임. 즉 분류의 대상을 결정해주는 것임. 학습이 되고 나면 직선이 생성되고, 2차원에서는 직선에 대한 방정식을 얻을 수 있음. 이를 수십 수천 번 학습하여 학습 된 이후 직선 위에 놓일 확률과 직선 아래에 놓일 확률을 통해 환율이 상승할지 하락할지 알 수 있음.

2차원에서의 매핑결과(예시)



자료: 유안타증권 리서치센터

3차원(고차원)으로 매핑결과(예시)



자료: 유안타증권 리서치센터



Part. 2

SVM 기반 환율예측 모델링

머신러닝 기반 환율예측 사용변수

- 미래의 환율수준 도출을 위하여 사용할 수 있는 변수는 결국 이론적으로 알려져 있듯이 환율과 밀접한 영향을 가지는 변수로 선정해야 함.
- 환율은 frequency가 매우 짧은 특징을 지니고 있으므로 데이터 활용 측면에서 다량 확보가 용이함.
- 이와 같은 특징을 지니며 환율에 밀접한 영향을 가지는 변수는 결국 금리인 것으로 판단됨. 또한 짧은 주기로 집계할 수 있는 데이터는 주가지수, 지수변동성, 원자재가격, 타국 환율 등이 있음.
- 따라서 본 연구에서 input data로 활용하는 변수의 category는 금리, 주가, 주가변동, 원자재, 환율로 구성됨.
- 금리로 활용되는 지표는 각국의 장기금리(10년물), 단기금리(2년물), 물가채금리 등이 있음.
- 원자재가격은 유가와 금 가격으로 구성됨. 주가는 주요국의 대표 지수를 사용하였으며, 변동성지수가 동시에 포함됨. 또한 환율은 달러대비 환율로 구성됨.

SVM을 통한 환율 예측에 수행되는 변수 선정과 구성

구분	변수명	구분	변수명
금리	미국 장기금리(10년물)	주가	미국 S&P500 주가
	미국 단기금리(2년물)		VIX 지수
	미국 물가채금리(10년물)		일본 NIKKEI255 주가
	미국 물가채금리(2년물)		NIKKEI255 변동성 지수
	일본 장기금리(10년물)		유로 STOXX50 주가
	일본 단기금리(2년물)		STOXX50 변동성 지수
	일본 물가채금리(10년물)		독일 DAX30 주가
	유로존 장기금리(10년물)		DAX30 변동성 지수
	유로존 단기금리(2년물)		영국 FTSE100 주가
	독일 장기금리(10년물)		FTSE100 변동성 지수
	독일 단기금리(2년물)		브라질 IBOV 주가
	독일 물가채금리(10년물)		호주 S&P/ASX 200
	독일 물가채금리(2년물)		중국 상해종합지수
	영국 장기금리(10년물)		홍콩 항생지수
	영국 단기금리(2년물)		항생 변동성 지수
	영국 물가채금리(10년물)		한국 KOSPI200 주가
	브라질 장기금리(10년물)		KOSPI200 변동성 지수
	브라질 단기금리(2년물)		한국 KOSDOQ 주가
	브라질 물가채금리(10년물)	환율	유로존 유로/달러 환율
	브라질 물가채금리(2년물)		일본 엔/달러 환율
	호주 장기금리(10년물)		중국 위안/달러 환율
	호주 단기금리(2년물)		영국 파운드/달러 환율
	호주 물가채금리(10년물)		호주 달러/달러 환율
	중국 장기금리(10년물)		브라질 헤알/달러 환율
	중국 단기금리(2년물)	원자재	한국 원/달러 환율
	한국 장기금리(10년물)		금 선물가격
	한국 단기금리(2년물)		WTI 선물가격


자료: 유안타증권 리서치센터


회귀분석 vs 서포트 벡터 머신


- 사실 SVM과 회귀분석의 가장 큰 차이점으로 볼 수 있는 것은, 회귀분석의 경우에는 모형 설정상 자유도의 문제에 의하여 관측치보다 많은 설명변수를 포함시킬 수 없음.
- 그러나 SVM을 통한 머신러닝 기법은 오히려 적은 관측치를 확보하고 많은 투입변수를 활용하여 학습을 시키는 것에 좀 더 용이하며 좋은 성과를 낼 수 있음.
- 따라서 확보된 투입변수를 확장하여 고차원의 분류방식을 선택해야 성과가 좋아진다는 부분이 회귀분석과의 차이점.
- 금리, 주가, 변동성, 환율 등의 변수를 회귀분석으로 추정한다면 많아야 10여개 내외의 설명변수가 포함되지만 SVM에서는 변수의 제한이 없음.
- 본 연구는 더욱 정확한 예측을 위하여 원계열의 lag 값을 투입하였음.


원계열 변수에 lag를 추가로 도입하여 변수 확장

	var 1	var 2	...	var 60	var 61	var 62	...	var 120	var 121	var 122	...	var 180	var 181	...	var 240
t+1	1	1	1	0											
t+2	1	1	0	0	1	1	1	0							
t+3	0	0	1	1	1	1	0	0	1	1	1	0			
t+4	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1
t+5	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0
t+6	1	1	1	0	1	1	0	0	0	0	0	0	0	0	1
t+7	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0
t+8	0	1	0	0	1	1	0	0	1	1	1	0	1	1	0
t+9	1	1	0	1	0	1	0	0	1	1	0	0	1	1	1
t+10	1	1	1	1	1	1	0	1	0	1	0	0	1	1	0
t+11	1	0	1	1	1	1	1	1	1	1	0	1	0	1	0
t+12	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1
t+13	1	1	0	0	1	1	1	0	1	0	1	1	1	1	1
t+14	0	0	1	1	1	1	0	0	1	1	1	0	1	1	1
t+15	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1
t+16	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0
t+17	1	1	1	0	1	1	0	0	0	0	0	0	0	0	1
t+18	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0
t+19	0	1	0	0	1	1	0	0	1	1	1	0	1	1	0
t+20	1	1	0	1	0	1	0	0	1	1	0	0	1	1	1
t+21	1	1	1	1	1	1	0	1	0	1	0	0	1	1	0
t+22	1	0	1	1	1	1	1	1	1	1	0	1	0	1	0


원계열


원계열
Lag 1


원계열
Lag 2


원계열
Lag 3

자료: 유안타증권 리서치센터

참고: 이론모형 – 서포트 벡터 공간 생성과 최적화 과정

SVM은 초평면에서 입력된 자료(\vec{u})가 +인지 -인지 결정함

2차원 공간위에 결정경계와 \vec{w} 가 있을때 \vec{w} 가 결정경계 위에 있다면 (+), \vec{w} 가 결정경계 아래에 있다면 (-)로 구분. 즉 분류규칙을 다음과 같이 설정

$$\vec{w} \cdot \vec{u} \geq c \quad \text{식(1)}$$

여기서 c 는 임의의 상수, 아직 \vec{w} 와 c 는 결정되지 않았지만 조건이 만족된다면 +라고 분류규칙을 정함. 식(1)을 변형시키면 식(2)와 같음

$$\vec{w} \cdot \vec{u} + b \geq 0 \quad \text{식(2)}$$

입력된 자료가 훈련될 때, (x_+) 과 (x_-) 로 훈련샘플을 구분할 수 있음. 선형 SVM에서는 (x_+) 과 (x_-) 는 가장 가까운 결정경계로부터 동일하게 δ 만큼 떨어져야 하거나 δ 보다 크게 떨어져야 함

$$\vec{w} \cdot \vec{x}_+ + b \geq \delta \quad \text{식(3)}$$

$$\vec{w} \cdot \vec{x}_- + b \leq -\delta \quad \text{식(4)}$$

이를 변수 y_i 로 정리하고 $y_i = \begin{cases} +1 & \text{for } x_+ \\ -1 & \text{for } x_- \end{cases}$ 식(3), (4)를 합치면, 조건에 따라 분류된 식(5)와 같은 결과를 산출할 수 있음

$$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 \quad \text{식(5)}$$

최적의 초평면을 찾는 최적화는 식(5)를 만족하면서 (x_+) 와 (x_-) 의 거리를 최대로 하는 w 와 b 를 찾는 것임. 결정경계는 초평면 위에 있는 두 경계 사이에 정확히 가운데 위치함.

두 경계 평면 사이의 거리는 $\frac{2}{\|\vec{w}\|}$ 로 최적화 조건은 식(6)과 같이 표현됨.

$$\min_{w,b} \frac{1}{2} \|\vec{w}\|^2 \quad \text{식(6)}$$

$$s. t. y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, \text{ for } i = 1 \dots m$$

훈련샘플을 완벽하게 구분해내는 평면이 없을 경우 추가적인 변수 $\xi_i \geq 0$ 을 도입하여 식(6)의 제약조건을 다음과 같이 수정할 수 있음.

$$y_i(\vec{w} \cdot \vec{x}_i + b) - \xi_i \geq 1, \text{ for } i = 1 \dots m$$

이때 목적함수를 수정해 구분하지 못하는 훈련샘플에 대해 제약을 하도록 하고 식(7)과 같이 최적화 조건을 수정할 수 있음.

$$\min_{w,b,\xi_1,\dots,\xi_m} \frac{1}{2} \|\vec{w}\|^2 + \sum_{i=1}^m C_i \xi_i$$

$$s. t. y_i(\vec{w} \cdot \vec{x}_i + b) - \xi_i \geq 1, \text{ for } i = 1 \dots m$$

$$\xi_i \geq 0, \text{ for } i = 1 \dots m \quad \text{식(7)}$$

참고: 이론모형 – 분석모형 선정

- 기본 추정식은 아래와 같음

$$krw\ dirctin^{t+1} = f(F^{t-4\cdots 0}_{Bond}, F^{t-4\cdots 0}_{Stock\ Price\ Index}, F^{t-4\cdots 0}_{Gold\ Future\ Price}, F^{t-4\cdots 0}_{Oil\ Future\ Price})$$

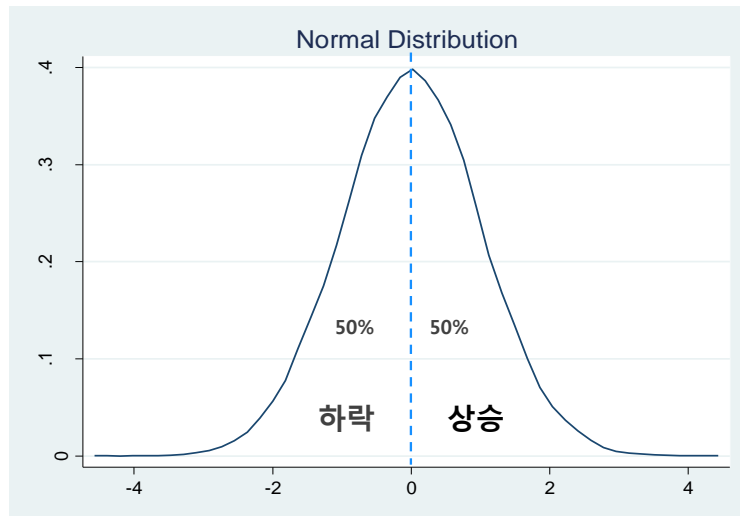
- $F^{t-4\cdots 0}_{Bond}$: t 기 부터 t-4 기까지 각 국가별 채권금리
 - $F^{t-4\cdots 0}_{Stock\ Price\ Index}$: t 기 부터 t-4 기까지 각 국가별 주가지수, 주가변동성
 - $F^{t-4\cdots 0}_{Gold\ Future\ Price}$: t 기 부터 t-4 기까지 금 선물가격
 - $F^{t-4\cdots 0}_{Oil\ Future\ Price}$: t 기 부터 t-4 기까지 유가(WTI) 선물가격
 - $F^{t-4\cdots 0}_{FRX}$: t 기 부터 t-4 기까지 각 국가의 화폐/달러 환율
- $krw\ dirctin^{t+1}$ 은 SVM 예측분류에 따라 도출된 환율의 t+1기 원/달러 환율의 움직임으로써 최종적으로 아래의 수식을 추정

$$krw\ dirctin^{t+1} = \begin{cases} 1, & \text{원달러 환율 상승} \\ 0, & \text{원달러 환율 하락} \end{cases}$$

자료의 변형

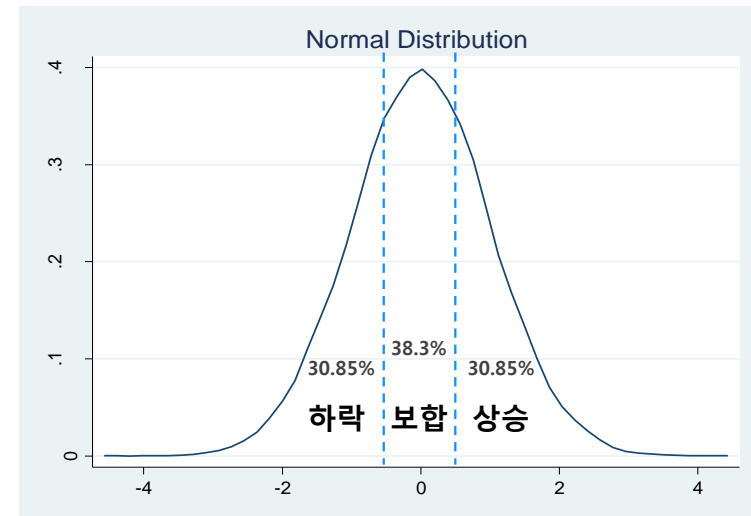
- 모형의 설명 부분에서도 언급되었듯이 SVM은 binary data 혹은 class로 구분된 데이터 등 discrete data를 cluster화 시킨다는 측면에서 의의를 지님. 따라서 환율을 추정함에 있어서도 level 값 혹은 변화율 값은 continuous 변수의 형태이기 때문에 분석에 적합하지 않음. 이와 같은 문제를 해결하기 위하여 변동을 단순화할 필요.
- 본 연구에서는 Case 1과 Case 2로 구분하여 분석을 진행. Case 1은 주간단위의 환율을 전주대비 Up과 Down의 두 국면으로 변환한 후 머신러닝을 통하여 다음주 환율이 현재수준 대비 높아질지 혹은 낮아질지에 대한 분석을 진행. Case 2는 환율의 전주대비 변화율의 분포가 거의 정규분포를 따르고 있으므로 평균을 중심으로 $\pm 0.5\sigma$ 이내에 위치하면 보합, 그 이상이면 상승, 그 이하이면 하락의 세 국면으로 구분하여 분석.

Case 1의 전주대비 환율변동은 정규분포에서 평균 이하이면 하락, 이상이면 상승



자료: 유안타증권 리서치센터

Case 2에서는 평균을 중심으로 $\pm 0.5\sigma$ 의 범위를 임계점으로 국면 구분

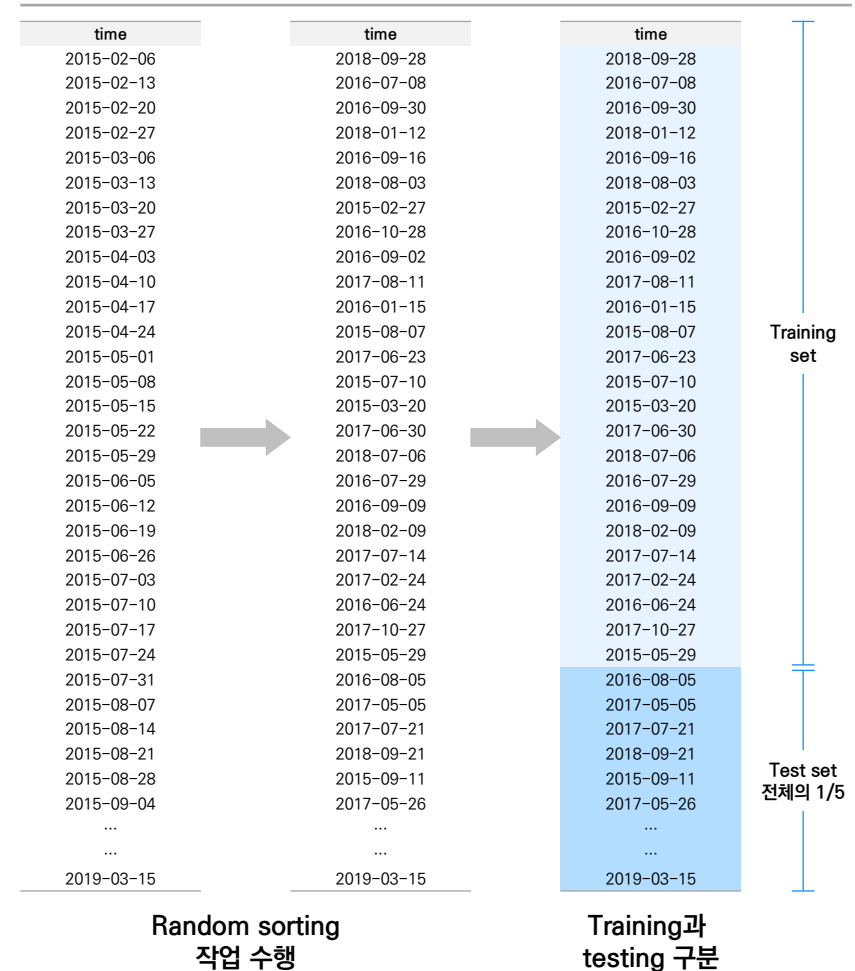


자료: 유안타증권 리서치센터

Training과 Test의 구분

- SVM은 주어진 정보를 활용하여 지도학습법을 진행하고, 이를 토대로 back-testing을 진행하는 방법. 따라서 전체 샘플을 training set과 test set으로 구분하는 작업이 필요함.
- 본 연구에서는 전체 225개의 샘플 중 4/5인 180개 샘플을 training set으로 선정하고, 나머지 45개의 샘플에 대해서 백테스트를 수행하도록 함.
- 다수의 SVM을 활용한 주가예측 등의 논문에서는 분석대상이 시계열 자료이기 때문에 training과 test를 구분할 때 과거의 일정 범위를 training set으로 정의하고 최근 수치를 test set으로 선정하는 경향.
- 그러나 모델의 적합도를 고려하는 동시에 한쪽으로 치우친 분포의 상황을 배제하기 위해서 본 연구에서는 random distribution 변수를 생성한 후 시계열에 상관없이 관측치를 shuffle하여 training과 test를 선정하였음.

Random sorting 작업 이후 training과 test set을 구분

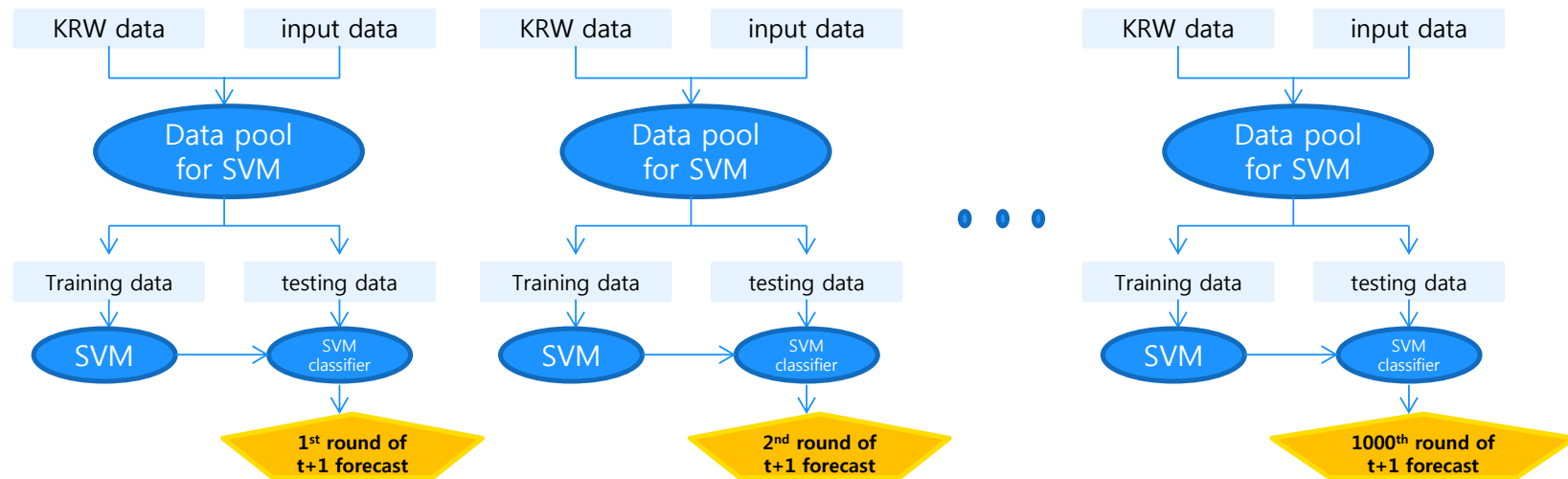


자료: 유안타증권 리서치센터

반복 시행(iteration n=1000)

- 사용된 변수의 training set과 test set을 random으로 선정한 또 다른 이유는 바로 반복시행에 있음. 아무리 머신러닝이라고 하더라도 단지 1~2회의 모델링 및 분석횟수를 가지고 적합성이 높다고 언급하기는 어려움. 따라서 본 연구에서는 무작위 샘플링을 통하여 set을 구분하고 학습과정을 거치고 백테스트를 하는 일련의 분석과정을 1000회를 진행. 시행 횟수가 늘어날수록 실제 hit ratio에 가까워질 확률이 높아질 수 있음.
- 본 연구는 비지도학습이 아니라 지도학습의 형태를 가지기 때문에 현재 관측할 수 없는 미래의 값에 대하여 가상의 값을 부여한 후 분석을 진행. 즉 case 1의 경우에는 환율이 전주대비 상승하였을 때의 값을 부여한 후 1000회의 iteration을 수행하였고, 동일하게 전주대비 하락하였을 경우의 값을 부여하고 동일한 1000회의 iteration이 진행. 총 2000회의 반복시행 결과를 확인할 수 있음(마찬가지로 case 2의 경우 총 3000회의 반복 시행이 이루어짐).

1000회에 걸친 반복 시행을 통하여 적합성을 높이는 작업을 수행



자료: 유안타증권 리서치센터



Part. 3

머신러닝을 통한 환율예측 결과

Case 1 분석결과 – 추정의 정확성 판단

- 추정의 정확성을 판단하기 위하여 iteration 회차별 추정값과 실제값의 일치성을 판단함. 기대확률 개념을 활용하여, 추정값이 하락했을 때 실제값이 하락으로 적절히 추정할 확률은 1000번의 시행시 평균으로 볼 때 64.9%를 기록. 상승으로 추정될 때 실제값이 상승인 경우는 66.1% 정도 정확성을 보임.
- 각 관측치별 실제값과 추정값의 정확성을 볼 때 하락 확률과 상승 확률의 차이는 크게 발생하지 않고 있지만 두 국면을 구분한다는 측면에서 볼 때 절반 이상의 정확성을 보인다는 측면에서 적절한 예측이 이루어짐.

회차별 추정의 정확성 판단

	Prob (실제하락 추정치하락)	Prob (실제상승 추정치상승)
1회차	72.0	80.0
2회차	61.5	68.4
3회차	76.9	53.1
4회차	66.7	62.9
5회차	42.3	73.7
6회차	65.0	40.0
...
...
999회차	60.0	68.0
1000회차	88.2	64.3
전체	64.9	66.1

자료: 유안타증권 리서치센터

각 관측치별 실제와 추정값의 괴리 및 해당 확률 추정 결과

obs	실제값	추정값	하락확률	상승확률
1	0	0	0.505	0.495
2	0	0	0.563	0.437
3	0	0	0.736	0.264
4	1	1	0.339	0.661
5	1	0	0.578	0.422
6	1	0	0.539	0.461
7	0	0	0.591	0.409
8	1	1	0.459	0.541
9	1	0	0.572	0.428
10	0	1	0.494	0.506
11	1	0	0.529	0.471
12	0	1	0.401	0.599
13	0	0	0.500	0.500
14	1	1	0.492	0.508
15	1	1	0.320	0.680
16	0	0	0.545	0.455
17	0	0	0.564	0.436
18	0	0	0.500	0.500
19	0	0	0.613	0.387
20	1	0	0.516	0.484
21	0	0	0.537	0.463
22	1	1	0.417	0.583
23	0	1	0.472	0.528
24	1	1	0.445	0.555
25	1	0	0.583	0.417
26	1	1	0.452	0.548
27	0	0	0.663	0.337
28	0	0	0.563	0.437
29	1	1	0.413	0.587
30	1	1	0.467	0.533
31	1	0	0.610	0.390
32	0	0	0.637	0.363
33	1	1	0.411	0.589
34	0	0	0.624	0.376
35	1	1	0.487	0.513

자료: 유안타증권 리서치센터

Case 1 분석결과 – 반복시행 결과와 확률

- 1000회의 반복시행을 통하여 추정된 전체 test data에 대한 적합성을 판단한 결과는 정확한 추정을 하는 경우가 65.6% 수준인 것으로 나타남. 즉 본 모형을 통하여 차주의 환율이 상승할지 혹은 하락할지에 대한 판단을 내릴 때의 정확성 수준은 일정 수준 담보되고 있는 것으로 볼 수 있음.
- 시점별로 볼 때 1000회의 반복시행 중 예측치가 test set에 포함된 경우는 총 218회가 있으며, 이 중 하락할 확률이 산출되는데, 이에 대한 매 회차별 확률의 평균치는 77.9%로 확인됨. 이는 머신러닝으로 여러 상황에 대한 판단을 한 결과 금주보다 환율이 하락할 확률이 더 높다는 것으로 해석됨. 다른 예를 들어 t-2기의 환율은 하락할 확률(29.6%)보다 상승할 확률(70.4%)이 더 높은 것으로 볼 수 있음.
- 결국 모형에 대한 신뢰도는 65.6% 정도가 되는 것으로 볼 수 있으며, 각 관측치별 상승과 하락에 대한 판정 역시 확률로 도출됨을 알 수 있음.

1000회의 반복시행으로 도출된 적합성 결과 hit ratio는 65.6%를 달성

65.6%의 hit ratio		예측값		Total
		하락	상승	
실제값	하락	13,115	8,409	21,524
	상승	7,076	16,400	23,476
Total		20,191	24,809	45,000

자료: 유안타증권 리서치센터

예측치에 대하여 상승을 기록할 확률과 하락을 기록할 확률을 구분하여 제시할 수 있음

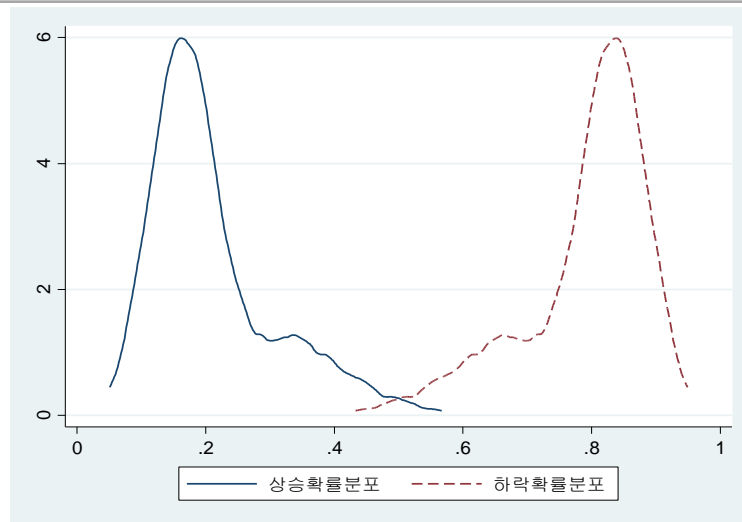
	Variable	Obs	Mean	SD	Min	Max
t+1예측치	Prob (하락)	218	0.779	0.107	0.477	0.942
	Prob (상승)	218	0.221	0.107	0.058	0.523
t기	Prob (하락)	217	0.779	0.105	0.477	0.941
	Prob (상승)	217	0.221	0.105	0.059	0.523
t-1기	Prob (하락)	191	0.422	0.068	0.245	0.611
	Prob (상승)	191	0.578	0.068	0.389	0.755
t-2기	Prob (하락)	211	0.296	0.068	0.111	0.459
	Prob (상승)	211	0.704	0.068	0.541	0.889
t-3기	Prob (하락)	230	0.680	0.072	0.500	0.879
	Prob (상승)	230	0.320	0.072	0.121	0.500

자료: 유안타증권 리서치센터

Case 1 분석결과 – 예측치 분포 추정

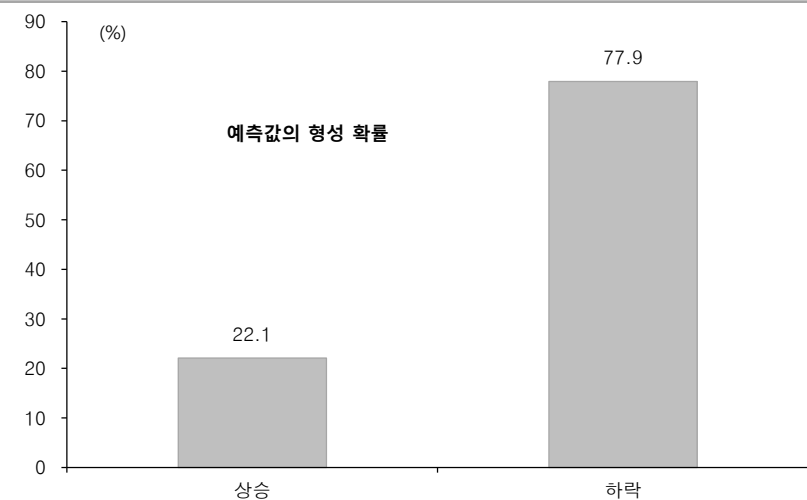
- 아래의 두 그림은 t+1기의 환율 예측치의 방향성에 대하여 주어진 확률이 어떻게 분포하고 있는지를 알려주는 그래프임. 분포표를 보면 상승확률분포는 중간값이 약 0.2 정도에 위치하고 있으며, 전반적으로 0.5보다 낮은 위치에 있는 것으로 볼 때, 상승할 확률은 상대적으로 낮은 것으로 해석. 반면 하락 확률은 중간값이 0.8 정도에 위치하고 있기 때문에 다음주의 환율 방향성이 하락 쪽으로 좀 더 가능성이 높은 상황인 것으로 판단.
- 이와 같은 분석 결과를 단순화하여 분포의 평균치를 제시한 값이 아래의 막대그래프임. t+1기의 환율 방향성에 대한 분포의 평균값으로 다음주의 환율이 상승할 가능성을 22.1%로 정의하고 하락할 가능성을 77.9%로 제시할 수 있음.

t+1기의 예측치가 상승할 확률과 하락할 확률에 대한 확률분포



자료: 유안타증권 리서치센터

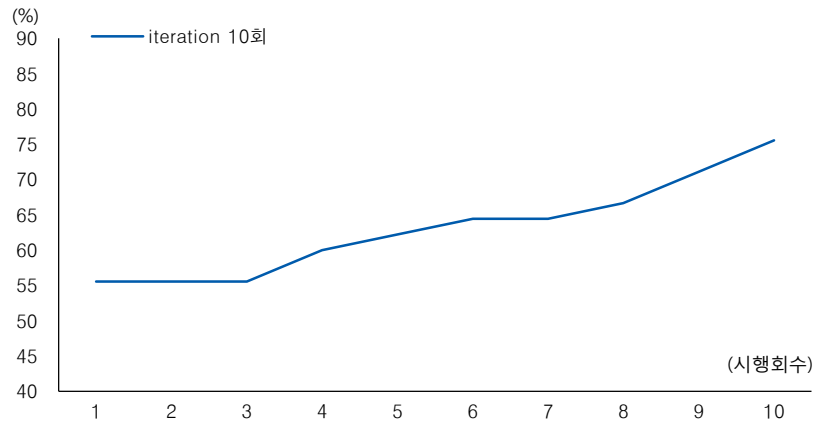
평균치로 보면 하락할 확률이 77.9%, 상승할 확률이 22.1%로 도출됨



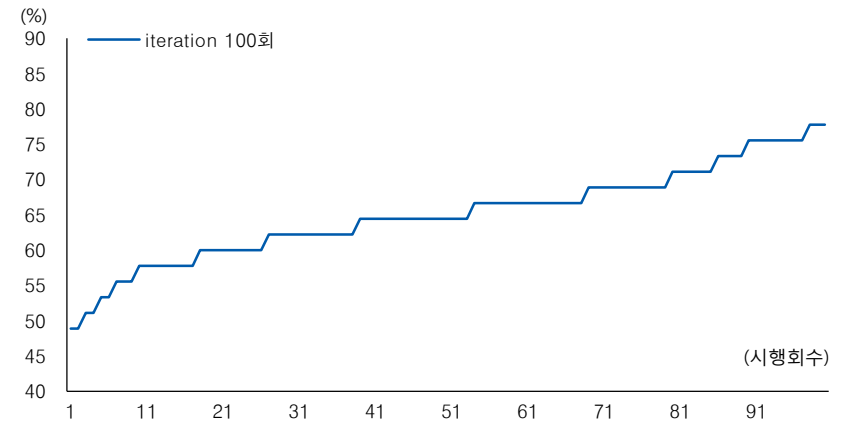
자료: 유안타증권 리서치센터

참고: 본 연구에서의 시행 회수별 hit ratio 비교

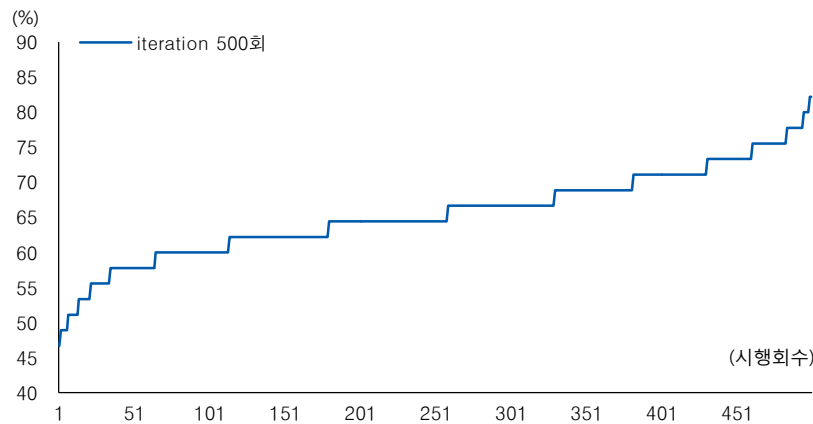
10회의 반복추정시 hit ratio 추이



100회의 반복추정시 hit ratio 추이

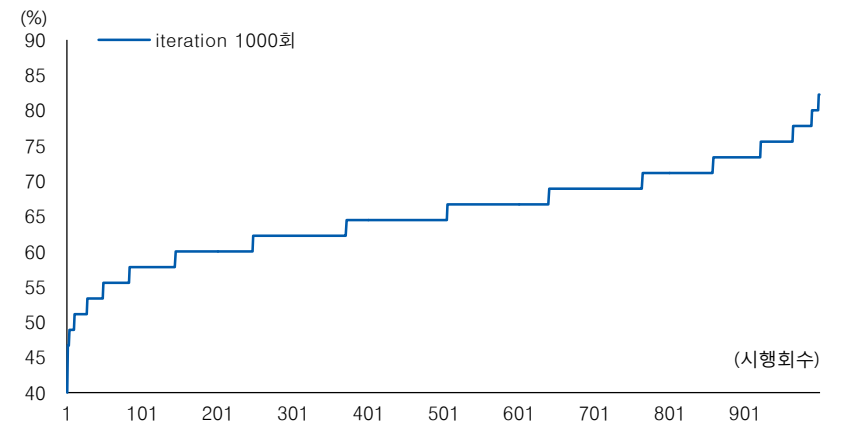


500회의 반복추정시 hit ratio 추이



자료: 유안타증권 리서치센터

1000회의 반복추정시 hit ratio 추이

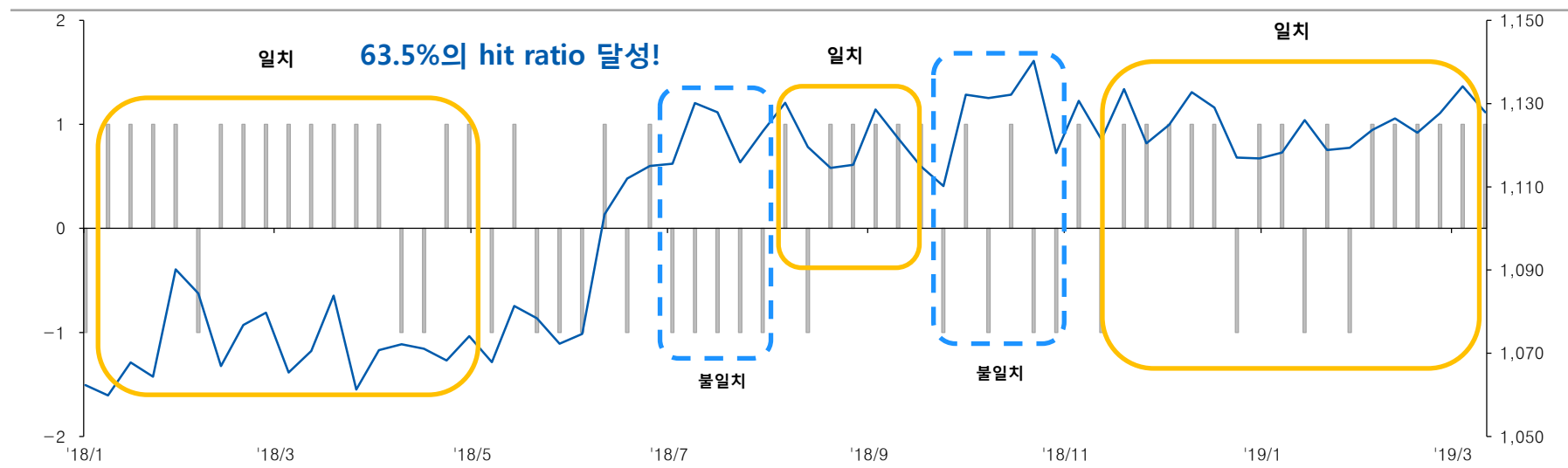


자료: 유안타증권 리서치센터

백테스트 - 실제 환율 움직임과의 비교

- 조금 더 정확한 백테스트를 위하여 2018년 1월 이후 현재까지의 환율수준을 상승과 하락 국면으로 구분하여 실제값과 추정값의 차이를 분석함. +1의 막대그래프를 가지는 경우는 정확한 추정이 이루어진 상황을 의미하며, -1의 막대그래프는 틀린 추정을 한 것으로 볼 수 있음. 전반적으로 63회의 추정 중 40회의 추정이 정확히 이루어졌기 때문에 약 63.5%의 정확도를 가지는 것으로 해석됨.
- 특이한 점은 환율이 안정적으로 횡보하는 국면에서 다음주의 환율 방향성을 예측할 때 정확도가 높은 것이 확인됨. 반면 structural break 등이 발생한 2018년 6월~7월, 2018년 10월~11월 등의 국면에서는 불일치하는 모습을 보임.
- 이러한 현상은 상대적으로 큰 변동이 없이 움직이는 안정적인 흐름에서는 머신러닝이 정확성을 높이는 반면, 국면이 변환되는 시점에서는 다른 경제변수 등의 고려가 필요한 상황이 발생하는 것으로 해석.

2018년 1월 이후 실제 원달러 환율의 움직임과 SVM으로 추정된 환율의 일치성 판단



주: 막대그래프의 +1은 실제값과 추정값이 일치한 시점을 의미하고, -1은 추정이 잘못된 상황을 의미함.

자료 : Bloomberg, 유안타증권 리서치센터

기존 연구와의 비교 - SVM을 활용한 KOSPI, 항생지수 예측

- Yanshan Wang, In-Chan Choi(2018)에서는 2002년부터 2011년까지의 약 10년간 항생지수의 구성 종목, KOSPI200 종목의 개별 주가, 환율 등을 수집해 주성분 분석(PCA)을 활용하여, 지수의 움직임에 영향을 미치는 고유치 이상의 값을 추출하여 서포트 벡터 머신(support vector machines)을 통한 주가의 방향성을 예측하였음.
- 주성분 분석(PCA)-SVM활용 방향성 예측결과 KOSPI 지수는 평균 61.76%의 예측정확도를 보이는 것으로 나타났으며, 항생지수는 평균 62.8%의 예측정확도를 보이는 것으로 나타남.

KOSPI 지수 예측 정확도

Iteration	PCA-SVM	SVM	PCA-ANN	ANN	RW
1	63.85%	63.85%	69.87%	75.90%	55.42%
2	55.87%	55.87%	56.68%	57.48%	47.36%
3	64.14%	59.34%	66.26%	63.82%	50.40%
4	59.34%	49.19%	56.45%	49.19%	52.82%
5	58.10%	58.89%	58.10%	59.68%	51.77%
6	69.75%	63.34%	59.36%	47.80%	51.39%
7	61.29%	61.29%	53.62%	60.88%	49.37%
Average	61.76%	58.82%	60.05%	59.25%	51.22%
Std	4.62%	5.05%	5.84%	9.45%	2.56%

자료 : Yanshan Wang, In-Chan Choi(2018), 유안타증권 리서치센터

항생 지수 예측 정확도

Iteration	PCA-SVM	SVM	PCA-ANN	ANN	RW
1	61.53%	59.10%	50.60%	45.74%	46.15%
2	59.51%	55.46%	53.03%	53.84%	45.74%
3	64.22%	56.50%	69.91%	56.09%	50.00%
4	57.14%	57.14%	68.16%	66.53%	44.89%
5	67.87%	53.41%	61.44%	59.43%	51.80%
6	61.84%	60.64%	66.66%	55.82%	52.61%
7	67.47%	66.66%	60.16%	56.50%	52.54%
Average	62.80%	58.42%	61.42%	56.28%	49.15%
Std	3.97%	4.33%	7.46%	6.22%	3.46%

자료 : Yanshan Wang, In-Chan Choi(2018), 유안타증권 리서치센터

기존 연구와의 비교 - SVM을 활용한 대만 증시 추세 예측

- Yuling LIN, Haixiang GUO and Jinglu HU(2013)에서는 2008년 7월 16일부터 2011년 11월 21일까지의 Taiwan 50 Index 구성 종목과 각 종목의 PER, ROI, OBV, RSI10, AR10, BR20 등 53개의 기술적 분석 자료를 학습자료로 활용해 SVM을 통한 대만증시의 추세에 대해 예측을 시도함.
- 예측결과 SVM 모형에서 Taiwan 50 Index의 예측 정확도는 최대 61.36%, 최소 56.82%로 나타났으며, 가중치 조정 모형에서는 최대 68.18%의 예측 정확도를 기록함.

비가중치 적용 투입 자료 학습 SVM 모형 분석결과

Model	HHC	TSC	EC	TW50
SVM_lin	63.64	63.64	65.91	61.36
	[28/44]	[28/44]	[29/44]	[27/44]
SVM_rbf	61.36	61.36	63.64	59.09
	[27/44]	[27/44]	[28/44]	[26/44]
SVM_poly	56.82	52.27	63.64	56.82
	[25/44]	[23/44]	[28/44]	[25/44]
SVM_slg	54.55	54.55	56.82	56.82
	[24/44]	[24/44]	[25/44]	[25/44]
SVM_ql	65.91	63.64	65.91	63.64
	[29/44]	[28/44]	[29/44]	[27/44]

자료 : Yuling LIN, Haixiang GUO and Jinglu HU(2013), 유안타증권 리서치센터

가중치 조정을 통한 자료 학습 SVM 모형 분석결과

Model	HHC	TSC	EC	TW50
C-SF+W SVM_lin	68.18	63.64	68.18	63.64
	[30/44]	[28/44]	[30/44]	[28/44]
C-SF+W SVM_rbf	70.45	68.18	68.18	68.18
	[31/44]	[30/44]	[30/44]	[30/44]
C-SF+W SVM_poly	68.18	63.64	68.18	63.64
	[30/44]	[28/44]	[30/44]	[28/44]
C-SF+W SVM_slg	68.18	63.64	65.91	63.64
	[30/44]	[28/44]	[29/44]	[28/44]
C-SF+W SVM_ql	75.00	68.18	70.45	68.18
	[28/44]	[30/44]	[31/44]	[30/44]

자료 : Yuling LIN, Haixiang GUO and Jinglu HU(2013), 유안타증권 리서치센터

기존 연구와의 비교 - SVM을 활용한 VKOSPI 예측 및 활용

- 김선웅 외(2016)에서는 ‘KOSPI 수익률’, ‘KOSPI 변동 폭’, ‘KOSPI 외국인 순매수 금액’, ‘외국인 선물 순매수 수량’, ‘개인 콜옵션 순매수 수량’, ‘개인 풋옵션 순매수 수량’, ‘VIX 전일종가 대비 당일 종가’ 를 학습 데이터로 투입하고, 서포트 벡터 머신(support vector machines)을 활용해서 KOSPI 변동성(VKOSPI)의 방향성 예측을 통한 옵션 매매와 벤치마크를 기반으로 한 옵션 매매의 수익률을 비교하였음.
- 분석결과 SVM을 기반으로 한 KOSPI 변동성(VKOSPI) 옵션매매의 결과가 벤치마크 기반으로 한 KOSPI 변동성(VKOSPI) 옵션매매의 결과보다 수익률이 높은 것으로 나타남.

SVM기반, VKOSPI의 방향성 예측을 통한 옵션 매매와 벤치마크 기반 옵션 매매 수익률을 비교

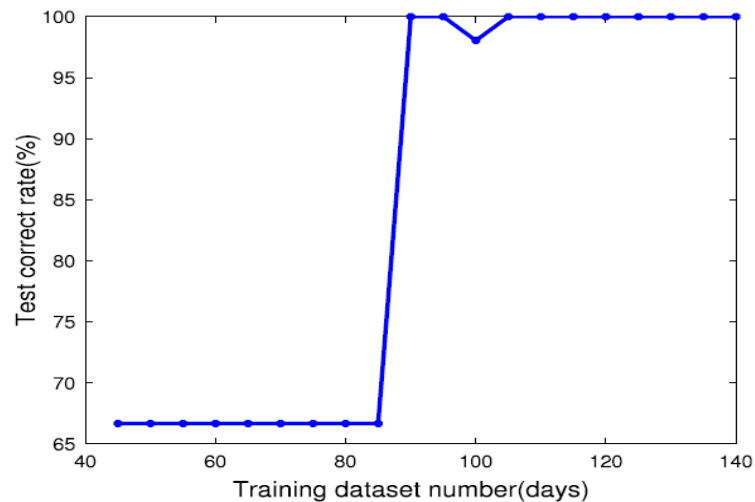
	Return	Number of transactions
SVM Strategy 1	-600,000	57
SVM Strategy 2	1,790,000	39
SVM Strategy 3	3,010,000	63
SVM Strategy 4	-1,860,000	44
SVM Strategy 5	-260,000	31
SVM Strategy 6	4,340,000	25
BM average	-5,085,000	100

자료 : 김선웅 외(2016), 유안타증권 리서치센터

기존 연구와의 비교 - SVM을 활용한 환율 방향성 예측

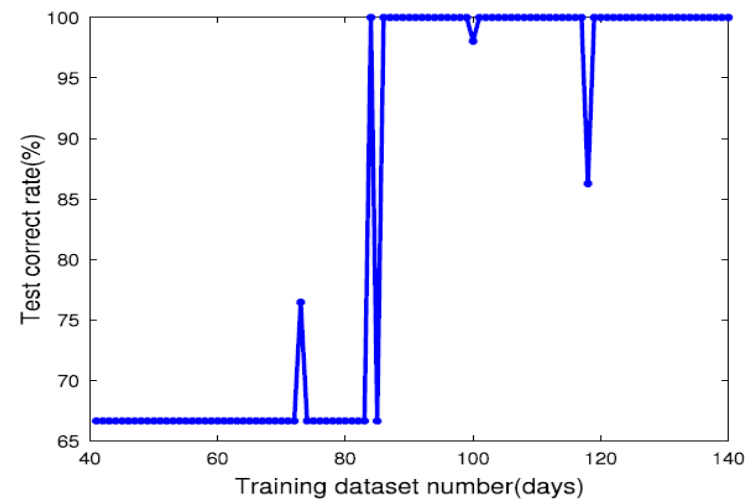
- Yubo Yuan(2012)에서는 2006년 부터 2012년까지의 위안화/달러 환율, 상하이 주가지수, 선전 주가지수 등을 학습자료로 구성하고, 위안화/달러 환율 방향성에 대한 예측을 위해, 다항식 커널 함수를 통한 SVM 모델을 활용하여 위안화/달러 방향성 예측을 시행하였음.
- 분석결과 다항식 커널 함수의 도입으로 학습량이 증가함에 따라 예측 정확도는 변화하는 모습을 확인 할 수 있으며, 평균 70% 이상의 예측 정확도를 기록하였음.

20회 학습결과



자료 : Yubo Yuan(2012), 유안타증권 리서치센터

100회 학습결과는 20회 학습결과에 비해 실제 hit ratio에 가까워질 확률이 높아짐



자료 : Yubo Yuan(2012), 유안타증권 리서치센터

Case 2 분석결과 – 추정의 정합성 판단

- 기대확률 개념을 활용하여, 추정값이 하락했을 때 실제값이 하락으로 적절히 추정할 확률은 1000번의 시행 평균으로 볼 때 46.3%를 기록. 보합 추정때 실제 보합일 확률은 35.5%, 상승 추정때 실제 상승은 48.5%의 정확성을 보임. 국면을 세개로 구분하면 산술적으로 33.3% 이상의 정합성이 확보되면 우수한 추정 결과로 볼 수 있음. 조건 충족 달성.
- 잘못 추정되는 경우라 할지라도 완벽히 반대방향의 실제값이 나타날 확률은 낮음. 최소한 보합 수준을 나타내고 있는 것으로 볼 때 완전히 다른 방향으로의 추정결과가 나타나는 경우는 제한적인 것으로 판단할 수 있음.

회차별 추정의 정합성 판단

	Prob(실제하락 추정치하락)	Prob(실제보합 추정치보합)	Prob(실제상승 추정치상승)
1회차	58.3	26.7	33.3
2회차	62.5	34.6	54.5
3회차	50.0	30.0	100.0
4회차	30.8	38.5	36.8
5회차	66.7	40.0	50.0
6회차	41.2	33.3	68.8
...
...
999회차	33.3	36.8	45.4
1000회차	100.0	35.7	66.7
전체	46.3	35.5	48.5

자료: 유안타증권 리서치센터

각 관측치별 실제와 추정값의 괴리 및 해당 확률 추정 결과

obs	실제값	추정값	하락확률	보합확률	상승확률
1	1	0	0.372	0.374	0.254
2	-1	-1	0.237	0.378	0.385
3	-1	-1	0.170	0.271	0.559
4	0	-1	0.135	0.306	0.558
5	1	0	0.365	0.423	0.212
6	-1	0	0.343	0.387	0.270
7	0	0	0.338	0.390	0.272
8	-1	0	0.328	0.398	0.275
9	-1	-1	0.240	0.348	0.412
10	-1	-1	0.161	0.309	0.531
11	-1	1	0.399	0.389	0.213
12	1	0	0.319	0.392	0.289
13	1	0	0.360	0.384	0.255
14	0	0	0.281	0.386	0.333
15	1	0	0.392	0.467	0.142
16	1	0	0.371	0.406	0.223
17	1	0	0.419	0.444	0.137
18	1	0	0.356	0.420	0.224
19	-1	0	0.331	0.372	0.297
20	-1	0	0.380	0.399	0.221
21	-1	-1	0.253	0.333	0.414
22	1	1	0.418	0.416	0.165
23	-1	-1	0.269	0.351	0.380
24	0	-1	0.174	0.350	0.475
25	-1	0	0.382	0.398	0.220
26	0	0	0.256	0.390	0.354
27	0	-1	0.236	0.351	0.413
28	0	0	0.380	0.426	0.194
29	1	0	0.403	0.431	0.167
30	-1	-1	0.147	0.327	0.526
31	0	1	0.431	0.422	0.147
32	1	0	0.378	0.450	0.172
33	0	0	0.300	0.398	0.302
34	1	0	0.392	0.419	0.189
35	0	0	0.342	0.374	0.284

자료: 유안타증권 리서치센터

Case 2 분석결과 – 반복시행 결과와 확률

- 1000회의 반복시행을 통하여 추정된 전체 test data에 대한 적합성을 판단한 결과는 정확한 추정을 하는 경우가 42.2% 수준인 것으로 나타남.
상대적으로 binary data를 분석한 Case1과 비교할 때 정확성은 낮아지고 있으나, 완벽히 반대의 추정 결과를 보이는 경우는 8.2%에 불과.
- 시점별로 볼 때 1000회의 반복시행 중 예측치가 test set에 포함된 경우는 총 220회가 있으며, 이 중 포함할 확률이 산출되는데, 이에 대한 매 회차별 확률의 평균치는 41.1%로 확인됨. 이는 전주대비 변화율의 분포가 거의 정규분포를 따르고 있으므로, 머신러닝으로 여러 상황에 대한 판단을 한 결과가 임계점 근처에서 크게 벗어나지 않아 예측정확도가 매우 우수한 것으로 판단할 수 있음.
- 다른 예를 들어 t-1기의 환율은 포함할 확률(34.2%) 상승할 확률(31.5%), 하락할 확률(34.3%)로 전주대비 변화율의 분포에 근접한 hit ratio를 보여주고 있음

1000회의 반복시행으로 도출된 적합성 결과 hit ratio는 42.2%를 달성

42.2%의 hit ratio		예측값			Total
		하락	보합	상승	
실제값	하락	4,382	6,665	2,205	13,252
	상승	1,485	6,344	7,455	15,284
Total		9,455	20,169	15,376	45,000

자료: 유안타증권 리서치센터

예측치에 대하여 상승을 기록할 확률과 하락을 기록할 확률을 구분하여 제시할 수 있음

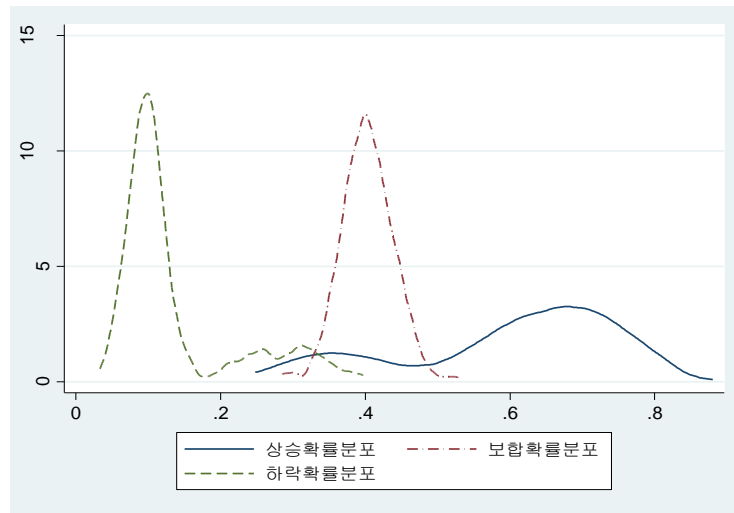
	Variable	Obs	Mean	SD	Min	Max
t+1예측치	Prob(하락)	220	0.279	0.044	0.181	0.367
	Prob(보합)	220	0.414	0.045	0.332	0.580
	Prob(상승)	220	0.307	0.049	0.164	0.412
t기	Prob(하락)	211	0.491	0.070	0.349	0.709
	Prob(보합)	211	0.347	0.060	0.203	0.476
	Prob(상승)	211	0.162	0.041	0.065	0.279
t-1기	Prob(하락)	197	0.343	0.054	0.205	0.497
	Prob(보합)	197	0.342	0.039	0.246	0.446
	Prob(상승)	197	0.315	0.054	0.161	0.464
t-2기	Prob(하락)	213	0.192	0.042	0.088	0.323
	Prob(보합)	213	0.472	0.056	0.345	0.647
	Prob(상승)	213	0.336	0.053	0.194	0.469

자료: 유안타증권 리서치센터

Case 2 분석결과 – 예측치 분포 추정

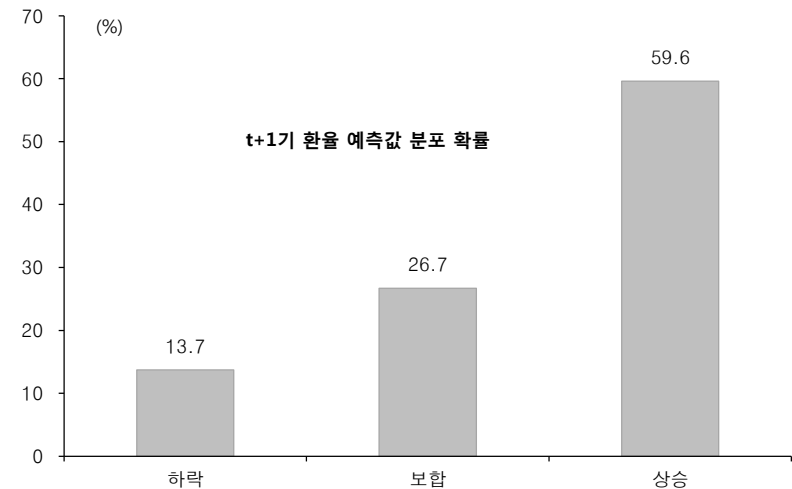
- 아래의 두 그림은 t+1기의 환율 예측치의 방향성에 대하여 주어진 확률이 어떻게 분포하고 있는지를 알려주는 그래프임. 분포표를 보면 하락 확률분포는 중간값이 약 0.1 정도에 위치하고 있으며, 전반적으로 0.2보다 낮은 위치에 있는 것으로 볼 때, 하락할 확률은 상대적으로 낮은 것으로 해석. 반면 상승 확률은 중간값이 0.7 정도에 위치하고 있기 때문에 다음주의 환율 방향성이 상승 쪽으로 좀 더 가능성이 높은 상황인 것으로 판단.
- 이와 같은 분석 결과를 단순화하여 분포의 평균치를 제시한 값이 아래의 막대그래프임. t+1기의 환율 방향성에 대한 분포의 평균값으로 다음주의 환율이 하락할 가능성을 13.7%로 정의하고 상승할 가능성을 59.6%로 제시할 수 있음.

t+1기의 예측치가 상승, 보합, 하락할 확률에 대한 확률분포



자료: 유안타증권 리서치센터

평균치로 보면 상승할 확률이 59.6%, 보합 26.7%, 하락할 확률이 13.7%로 도출됨



자료: 유안타증권 리서치센터



Part. 4

함의와 응용, 그리고 보완과제

이론, 이슈, 흐름이 모두 포함된 분석

- 우리가 추정하는 환율이라는 변수는 금리와 주가 등 각종 금융시장에서의 가격변수에 영향을 받게 됨. 지난 시점에서의 가격정보는 가격결정이라는 이론적 배경과 이슈, 흐름이 모두 포함되어 반영. 머신러닝은 기본적으로 학습을 필요로 하기 때문에 이슈, 흐름이 모두 포함되어 있는 환율에 적용할 수 있으며, 나아가 다른 가격 변수에 대한 머신러닝의 적용 자체가 용이함.
- SVM을 통해 가격변수와 예측하고자 하는 변수의 관련 여부를 분별할 수 있는 도구로서 확장할 수 있음. 즉, 과거에 이벤트로 인한 움직임이 반영된 가격변수에 포함된 정량화 할 수 없는 정보가 미래의 가격변동과 관련성 있다면, SVM을 통해 그 관련성과 방향성을 예측할 수 있을 것임. 직관적으로 시장의 효율성과 움직임을 충분히 반영하지 못하는 고전적인 분석에 의존하기 보다는 이러한 통계적인 기법을 활용하여 의사결정과 전략수립의 지표로 활용할 수 있음.

머신러닝의 응용과 확장을 통한 가격변수(주가, 환율) 움직임 예측 동향

	분석자료	분석머신	성능
J. Patel et al (2015)	CNX Nifty	SVM	머신러닝의 예측정확도가 5-7% 이상 높음
	BSE Sensex	Random Foresr	
R. Dash et al (2016)	BSE Sensex	CEFLANN	수익률 BSE: 47.2% S&P500: 24.28%
	S&P500		
L.D. Persio & O. Honchar (2016)	S&P500	Wavelet CNN	MSE: 0.2491 예측정확도: 54%
	EUR/USD		
M. Qinet al (2016)	Nikkei 225	ANN	MSE: 0.0043

자료: 유안타증권 리서치센터

금융산업의 머신러닝 도입은 거스를 수 없는 트렌드

기업명	활용	기업명	활용
Vanguard	트레이딩	요코하마은행	수익성 예측
Charles Schwab	운용정보	홍콩 증권선물위원회	업무 알고리즘
AIDYIA	투자이사 결정	Rebellion Reseach	수익률 분석, 예측
Royal Bank of Scotland	투자정보 제공	싱가포르 통화청	정책변수 검증

자료: 금융보안원(2017), 유안타증권 리서치센터

환율방향성 추정의 머신러닝 적용 성과와 함의

1. (트렌드) 환율예측을 머신러닝에 적용

기술의 진보에 따라 머신러닝을 적용한 분야는 기하급수적으로 확대되고 있음. 이러한 트렌드는 금융시장에서도 예외가 될 수 없는 상황으로 이와 같은 상황에 대한 대응 차원에서 본 연구에서는 머신러닝 기법을 환율 예측에 적용함.

2. (차별성) 반복시행 회수 증가를 통한 정합성 확보

본 연구는 머신러닝의 기능과 다양한 예측결과를 산출하기 위해, 기존의 여러 논문과 이슈자료에서 시도된 반복학습 회차보다 많은 1000회 반복시행을 통해 머신러닝의 한계를 확인하고, 실제 hit ratio에 가까워질 확률을 산출한 데에 의의가 있음.

3. (성과) 환율 방향성 예측 hit ratio 65.6% 달성 모형 개발

실제 본 연구를 통해 도달한 65.6%의 hit ratio는 차주의 환율이 상승할지 혹은 하락할지에 대한 판단을 내릴 때의 정확성 수준은 일정 수준 담보되고 있는 것으로 볼 수 있음.

4. (발견과 함의) 추정값 일치구간과 불일치구간의 특징적 현상 도출

(발견) 연구를 통해 관찰한 특이한 점으로, 환율이 안정적으로 횡보하는 국면에서 다음주의 환율 방향성을 예측할 때 정확도가 높은 것이 확인됨. 반면 structural break 등이 발생한 2018년 6월~7월, 2018년 10월~11월 등의 국면에서는 불일치하는 모습을 보임.

(함의) 이러한 현상은 상대적으로 큰 변동이 없이 움직이는 안정적인 흐름에서는 기계학습이 정확성을 높이는 반면 국면이 변환되는 시점에서는 다른 경제변수 등의 고려가 필요한 상황이 발생하는 것으로 해석될 수 있음.

머신러닝 활용 금융지표 예측의 응용과 보완 과제

5. (응용) 금융지표의 국면전환 검증을 위한 스크리닝

SVM 기법은 패턴을 분류하는데, 좋은 성과를 보이고 이러한 점을 착안하여 스크리닝 기능으로 활용한다면, 급격한 불일치 현상이 발생할 때 이러한 국면에 대한 예측 방법으로, SVM의 패턴분류 기능을 응용하여 structural break 검증을 위한 스크리닝 측면에서 활용할 수 있음.

6. (한계) 원인과 결과 사이의 Black Box

SVM의 불가피한 단점 중 하나는 분류 작업을 가상의 고차원 공간에서 수행하기 때문에 입력 변수와 이에 대한 결과값에 대해 직관적인 설명을 도출하기가 쉽지 않음. 모형이 환율 상승을 예측한 경우, 어떠한 변수가 크게 작용을 했는지, 혹은 변수가 어떠한 방향으로 얼마만큼 변하면 결과치가 다르게 나올지 예측과 유의성을 검증이 쉽지 않음.

7. (활용) 가격변수, 거시경제변수 등의 추세 및 방향성 판별

SVM과 같은 머신러닝의 분류 기능을 이용해 가격변수에 영향을 미치는 이슈, 거시경제흐름 등을 추세를 판별할 수도 있을 것이며 이러한 판별을 바탕으로 방향성을 예측하고 전략을 구성하는 데 유의미한 지표로 활용할 수 있음.

8. (결론 및 보완) 머신러닝의 가능성 확인, 매주 업데이트 수치 제공

본 연구를 통해 금융시장에서의 머신러닝 기법 도입 가능성을 확인. 지속적인 모델링과 커널기법, 추정기법 개선이 수행될 것이며, 추정된 환율예측 결과는 필자의 주간 경제노트 자료에 매주 업데이트하여 결과를 제시할 예정



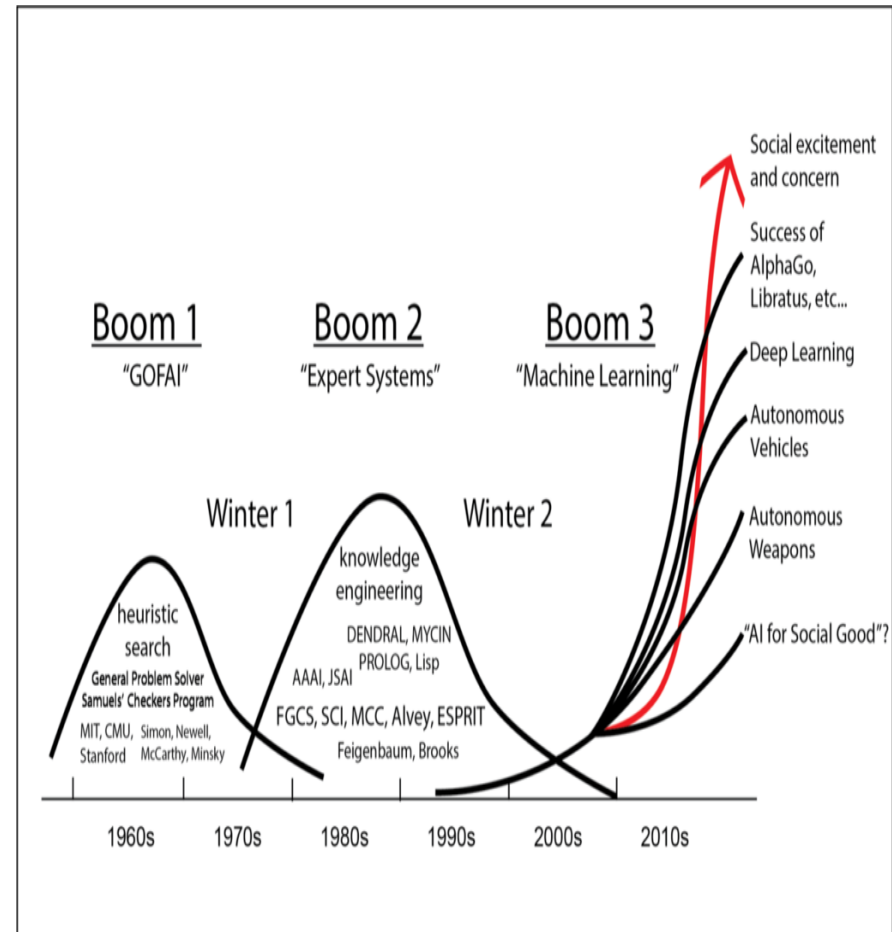
Appendix

머신러닝의 소개

머신러닝 개괄

- 러다이트 운동과 같이 기술 혁신의 과정에 적응한 자본가와 그렇지 못한 노동자 계층에서 반발하여 사회운동으로 나타난 이야기는 아주 유명한 사례.
- 머신러닝은 여러 산업과 학문에서 혁신을 일으키고 있음. 거시경제와 금융 분야 역시 예외는 아님.
- 본 연구는 머신러닝 기술들에 대한 전반적인 이론에 대해 간단히 알아보고, 본 연구에서 활용한 SVM(서포트벡터머신)의 학습방법에 대해 알아보았음.
- 이후 실제로 환율을 예측하고, 마지막으로 거시경제와 금융 분야의 다른 가격변수의 예측으로 확장성을 검토하였음.

인공지능의 발전사, 제3차 AI의 봄

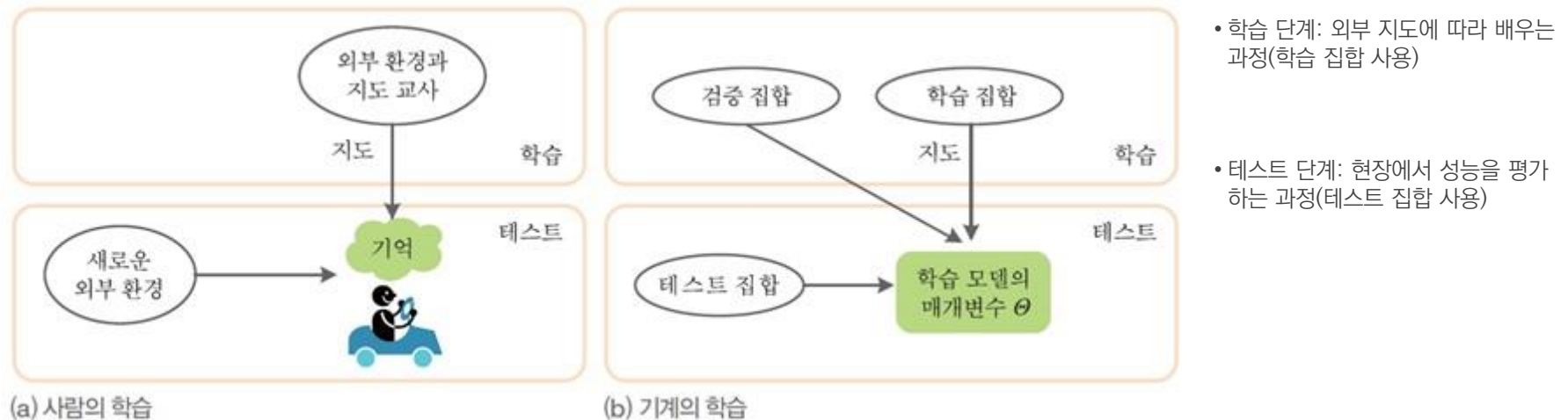


자료 : 松尾豊(2015), 人工知能は人間を超えるか: ディープラーニングの先にあるもの, Colin Garvey(2018)

머신러닝의 도입

- 기존의 통계적 알고리즘은 연구자가 수작업으로 매개변수를 설정하거나 규칙을 정하여 프로그램을 업데이트함. 반면 머신러닝은 학습 집합을 판단하여 필요한 최적의 설정을 자동으로 수행함.
- 머신러닝은 주어진 자료의 클래스를 구분하는 패턴인식(pattern classification)과 연속적인 값을 추정해야 하는 함수 근사화(function approximation) 문제를 수행할 수 있음.
- 시스템이 다루어야 할 환경이 바뀌면 새로운 환경에서 수집한 학습 집합으로 다시 학습하여, 변동성이 큰 대상을 예측하는데 유용.

이성적 학습과 머신(기계)의 학습 과정

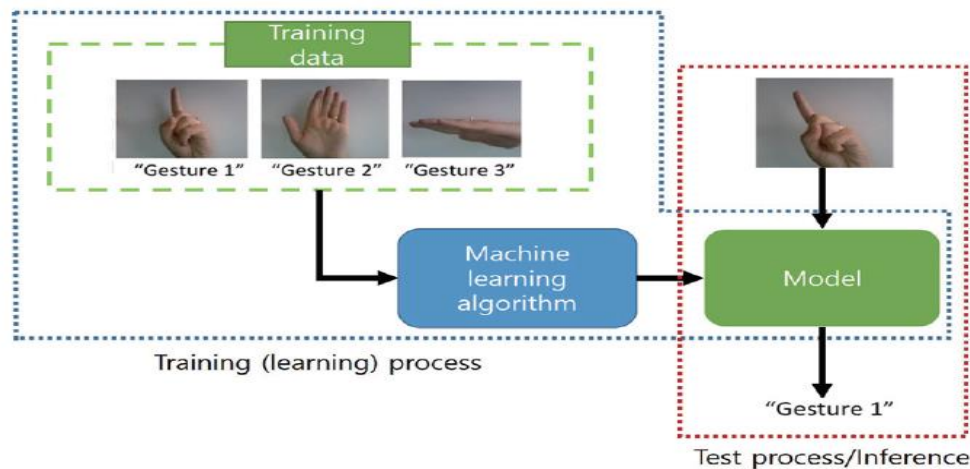


자료 : KIEE(2018), 유안타증권 리서치센터

지도학습(supervised learning)

- 지도학습(supervised learning)은 입력과 그 입력이 들어갔을 때 나와야 하는 출력의 예시를 통해 입출력관계를 학습하는 방법.
- 지도 학습의 가장 큰 특징은 Input 데이터(Traing set)에 레이블(label)이라는 답이 포함. 'Input'과 'Output'인 레이블이 세트가 된 훈련 데이터를 미리 준비하고, 어떤 입력이 들어오면 올바른 답이 나오도록 컴퓨터를 학습시키는 방법.
- 대표적인 지도 학습 알고리즘으로는 k-최근접 이웃(k-Nearest Neighbors), 선형 회귀(Linear Regression), 로지스틱 회귀(Logistic Regression), 결정 나무(Decision Tree)와 랜덤 포레스트(Random Forests), 신경회로망(Neural networks), 본 연구에서 활용할 서포트 벡터 머신(Support Vector Machines(SVM)) 등 이 있음.

기계의 지도학습 과정 예시



- 손을 촬영한 영상이 입력, 손의 자세에 대한 라벨을 출력 (트레이닝셋에 포함되지 않은)
- 학습이 완료되어 얻게 된 모델을 사용 새로운 입력에 대해서 라벨을 예측 (왼쪽 그림에서 붉은 색 점선 내부)

비지도학습(unsupervised learning)

- 비지도학습은 구체적으로 3가지 일을 수행할 수 있음.

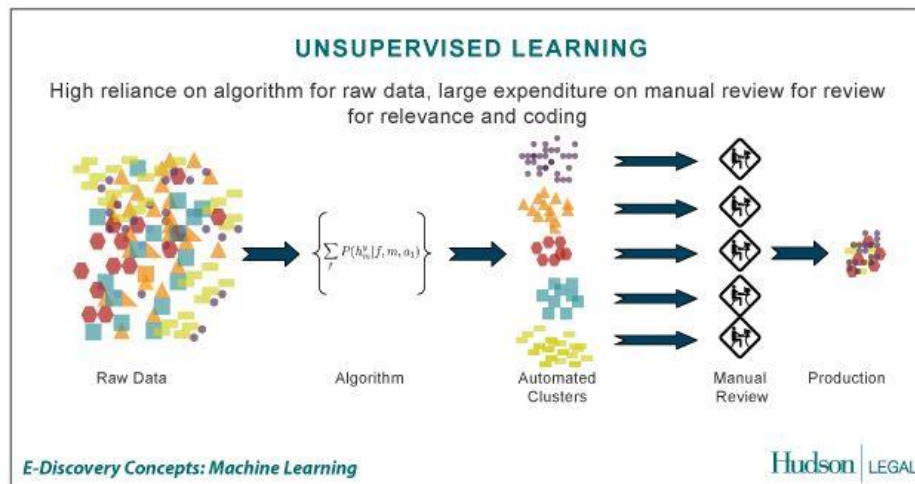
군집화(clustering): 데이터를 유사도에 따라서 여러 개의 그룹으로 나누는 작업

차원축소(dimension reduction): 특징학습으로 이는 데이터의 중요한 구조를 발견할 수 있으며 이후 데이터 처리에 편의성을 제공

생성모델학습(generative model learning): 데이터로부터 그와 유사한 데이터를 추가적으로 생성할 수 있는 확률 모델을 학습하는 방법

- 지도 학습과 달리, 비지도 학습의 자료에는 레이블(label)이라는 답이 존재하지 않음. 대신, 비지도 학습은 정답이 없이 목표만 주어지는 학습법으로 데이터 속에 있는 일정한 패턴이나 룰을 추출하는 것이 목적.

기계의 지도학습 과정 예시

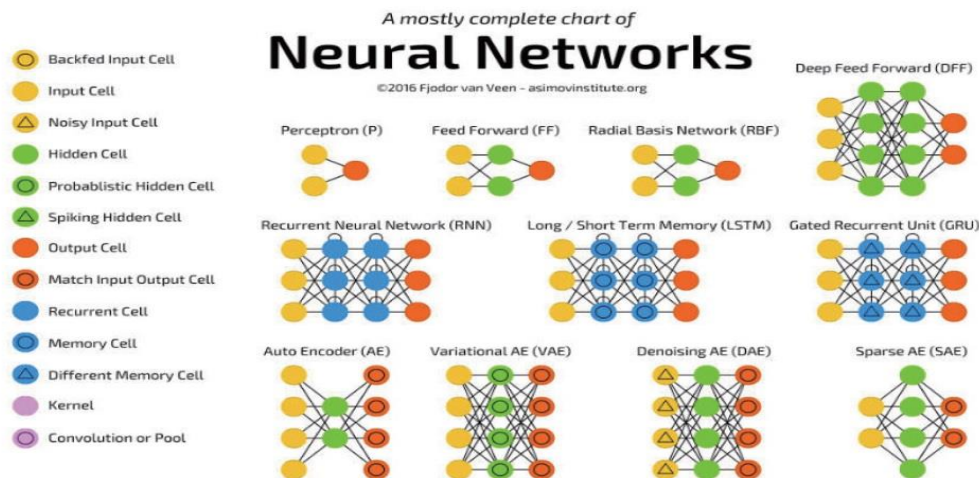


- 유사한 특징 벡터들을 끼리끼리 모으는 군집화 수행
- 차원축소 알고리즘을 통해 데이터 벡터의 차원을 축소
- 축소된 차원의 군집에서 유용한 정보 추출

인공신경망 학습

- 뇌는 뉴런이라는 단위로 구성되어 있음. 이러한 생물학적인 모델에서 영감을 받아 개별 뉴런을 수학적으로 모델링하고, 인공 뉴런 다수를 연결하여 인공신경망을 구성하여 기계학습에 활용.
- 구성된 인공신경망에는 많은 수의 파라미터(시냅스의 가중치)가 포함되어 있는 데 이러한 파라미터는 역전파 알고리즘(back-propagation algorithm)을 이용하여 찾음.
- 역전파 알고리즘을 통해서 학습된 신경망은 새로운 데이터를 처리하는 데 사용되며 이 과정은 추론(inference)이라 함.

다양한 신경망 학습 분류



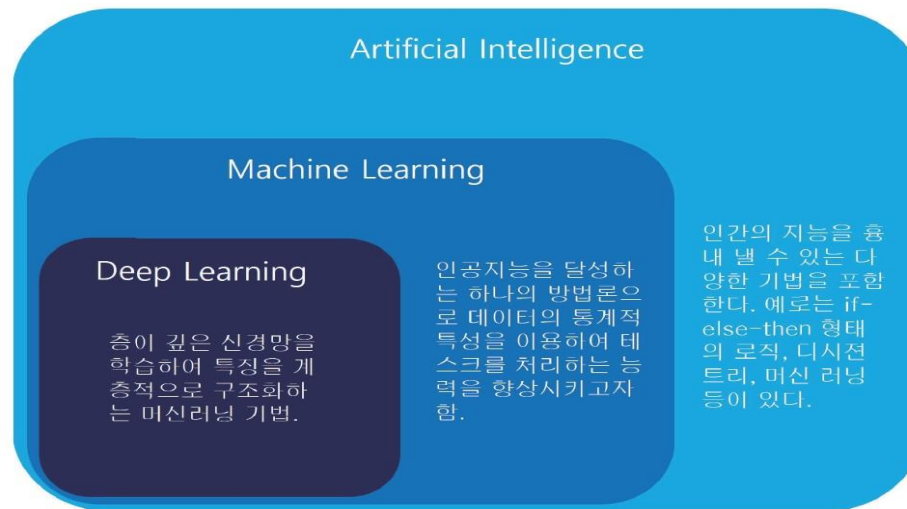
- 다양한 학습을 처리할 수 있도록 인공신경망 구조를 설계할 수 있음
- Feed forward 구조를 적용한 지도학습이 가장 많이 사용됨. 이 밖에 차원축소, 생성모델, 강화학습 등에 사용되는 모델을 구성할 수 있음
- 인공신경망 학습은 다양한 형태의 구조로 표현이 가능하며, 인공뉴런의 유연성(flexibility)의 특징에 기인함

자료 : KIEE(2018), 유안타증권 리서치센터

딥러닝

- 딥러닝은 규모가 큰 신경망을 사용했다는 점과 방대한 양의 데이터로 학습한다는 특징을 가지고 있음.
- 딥러닝 기반의 접근법은 데이터가 공급됨에 따라 성능이 꾸준히 향상되는 특성이 강하게 나타남.
- 딥러닝의 또 다른 특징은 end-to-end 학습이 가능함.
- 딥러닝은 시스템 설계과정에서 엔지니어가 개입하여 경험이나 직관에 의존하여 설계해야 하는 전통적인 접근과 다르게, 입력과 출력의 예시만으로 전체 시스템의 학습을 가능.

딥러닝, 기계학습, 인공지능 비교



- 최근 데이터의 증가, 병렬처리 하드웨어의 발달, 학습 알고리즘의 발전으로 매우 깊은 신경망의 학습이 가능
- 특히 “Deeper is better”라는 표현이 있을 정도로 딥러닝은 기계학습의 대세가 되고 있음

자료 : KIEE(2018), 유안타증권 리서치센터



이 자료에 게재된 내용들은 본인의 의견을 정확하게 반영하고 있으며 타인의 부당한 압력이나 간섭 없이 작성되었음을 확인함. (작성자: 정원일)

당사는 동 자료를 전문투자자 및 제 3자에게 사전 제공한 사실이 없습니다.

본 자료는 투자자의 투자를 권유할 목적으로 작성된 것이 아니라, 투자자의 투자판단에 참고가 되는 정보제공을 목적으로 작성된 참고자료입니다. 본 자료는 금융투자분석사가 신뢰할 만하다고 판단되는 자료와 정보에 의거하여 만들어진 것이지만, 당사와 금융투자분석사가 그 정확성이나 완전성을 보장할 수는 없습니다. 따라서, 본 자료를 참고한 투자자의 투자 의사결정은 전적으로 투자자 자신의 판단과 책임하에 이루어져야 하며, 당사는 본 자료의 내용에 의거하여 행해진 일체의 투자행위 결과에 대하여 어떠한 책임도 지지 않습니다. 또한, 본 자료는 당사 투자자에게만 제공되는 자료로 당사의 동의 없이 본 자료를 무단으로 복제 전송 인용 배포하는 행위는 법으로 금지되어 있습니다.

인공지능 | 머신러닝 거시경제를 예측하다

SVM 회귀분석을 통한 모델링