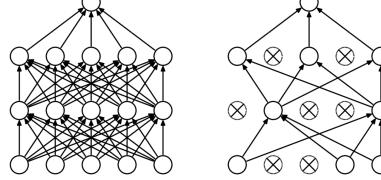
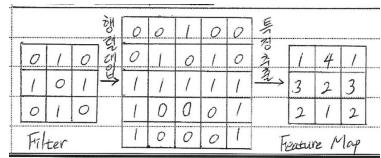
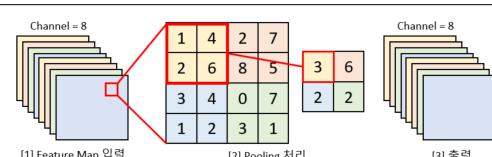
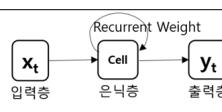
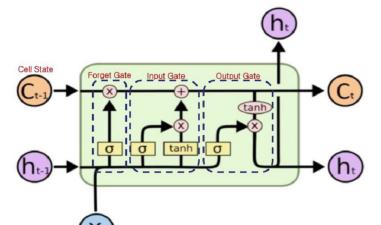
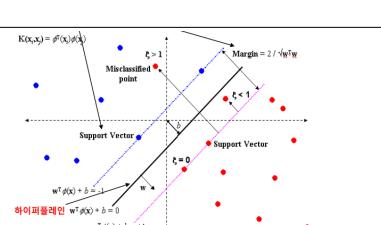
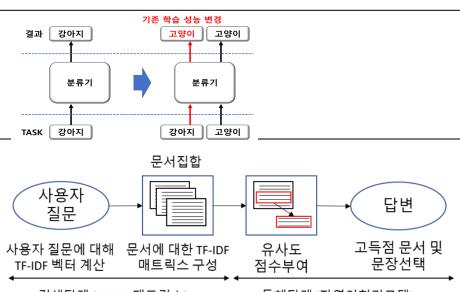
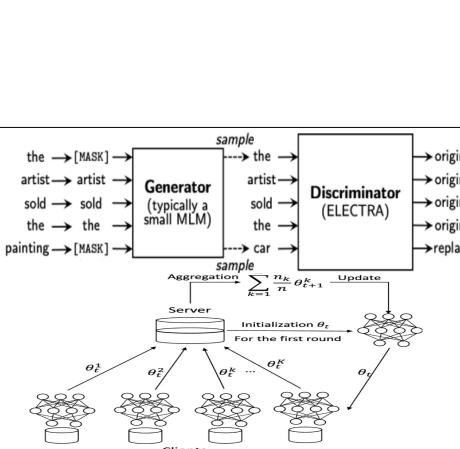
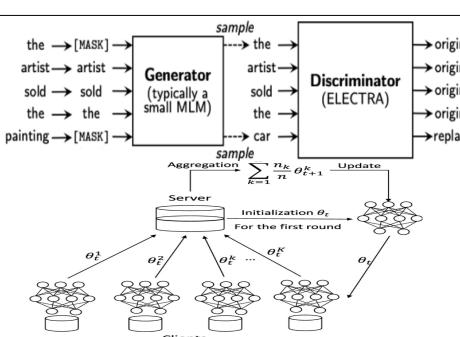
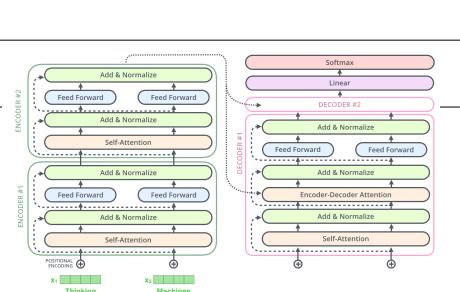
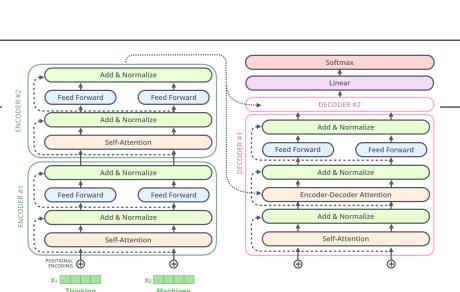


NO	포지	종포지	키워드, 두음	기본	서브노트	구성 및 추가정보
1	기계학습분류	감독학습/지도학습 (Supervised Learning)	분류 스케일 선선	○	<p>훈련데이터(Training Data)로부터 하나의 학습을 유도해내기 위한 기계학습(Machine Learning)의 한 방법</p> <p>ex) 고양이의 모습에 대한 득점을 학습시키고 유튜브에서 고양이 영상을 찾는 경우, 스크립에 일</p> <p>[알고리즘]</p> <ul style="list-style-type: none"> -분류(Classification): 주어진 데이터를 친합을 정해진 몇개의 클래스로 분류 >SVM(Support Vector Machines): 경계선 극점데이터 분석 >K-NN(K-Nearest Neighbor): 거리기반 많은 속성을 가진 클러스터 할당 >Convolutional Neural Network: Convolution, Pooling layer, Fully connected layer >RNN(Recurrent): 흐름에 정보자체 활용 -회귀(Regression): 데이터를 통해 주어진 입력에 대한 출력을 예측하는 방식 >선형회귀분석: 종속변수와 독립변수 관계 모델링 >신경망(Neural networks) 분석 <p>[평가]</p> <ul style="list-style-type: none"> -교차검증(Cross Validation) -정밀도(Precision)와 재현율(Recall) 측정 	
2	기계학습분류	무감독학습	군집 미니맥스 피카 적간	○	<p>일력값에 대한 특징이 주어지지 않는 기계학습</p> <p>[군집화 Clustering]</p> <ul style="list-style-type: none"> -K-means: K개의 데이터를 K개의 군집으로 분류하기 위해 거리 기반 반복 계산 >중심점이 변하지 않을 때까지 계속 반복, K개수 설정>초기화점, 군집, 중심점 설정, 재할당, 반복 -DBSCAN: 중심이동, 밀도기반, 밀도-EM Clustering: 규정분포 기반 E-Step, M-Step 반복, 군집화, 최적파라미터 찾기 -SOM: 자기조직 학습, 차원축소, 클러스터링 동시수행 <p>[차원축소] Reduce Dimension</p> <ul style="list-style-type: none"> -PCA(Principal Component Analysis): 주성분분석, 3차원 > 2차원 -ICA(Independent Component Analysis): 주어진 득점에서 서로운 득점 추출 <p>[회귀학습]</p> <ul style="list-style-type: none"> -GAN: 생성자(Generator)와 판별자(Discriminator)가 서로 경쟁하는 과정을 통해 정보를 학습 -DCGAN: Deep Convolutional GAN 	
3	기계학습분류	강화학습 (Reinforcement Learning)	Reward, 의사결정 유입 학습	○	<p>어떤 환경 안에서 장의원 예전트가 현재의 상태를 인식하여, 선택 가능한 행동들 중 보상을 최대화하는 행동 혹은 행동 순서를 선택</p> <p>[목표]</p> <ul style="list-style-type: none"> -Reward -수집적 의사결정: 관측순서에 따른 영향 <p>[알고리즘]</p> <ul style="list-style-type: none"> -Q-Learning: state/action > Q Table > Q-Value(Reward) -Deep Q-Learning: state > Q Network(신경망) > 여러 개 Q Value <p>[역회피학습(Inverse Reinforcement Learning)]</p> <ul style="list-style-type: none"> -보상수值得 구하는 학습, 회피 선택을 할 경향을 강화하여 계산 -보상수值得 계산: 보상수值得에 따라 보상 수值得 구하는 것은 매우 복잡한 문제 -보상수다중성: 보상 수值得 단일 속성이 아니라 다중성으로 설정되는 경우가 대부분 -미지로장출기기: 보상 수值得을 정의할 때 미지로 보상 수值得까지 초기적으로 고려 	
4	지도학습	딥러닝	다층신경망구조 심사준	○	<p>다양한 데이터가 있는 원데이터로부터 데이터 부분들의 관계를 통해 레이블을 자동으로 생성하여, 지도학습을 이용하는 학습 방법</p> <ul style="list-style-type: none"> -첨두가 미처 사람처럼 스스로 학습할 수 있는 인공지능 기술 -고장이 있을 때 이를 알려주고 알고양이 문제해결 <p>[동작배경]</p> <ul style="list-style-type: none"> -비지도학습기반 전처리기술 발전: 군집화기반, 노이즈 감소 -CNN(신경망): 가중치, 편향, Dropout, RNN(신경망: 신경인식, 자연언어), 하드웨어개선(GPU) <p>[알고리즘]</p> <ul style="list-style-type: none"> -DNN(신경망): 어려운 허트레이너(다중 퍼셉트론) -CNN(신경망): 대비 + 차원축소 -RNN(신경망): 신경망 연결의 순환구조, 허트레이너값 저장 후 이용 	
5	지도학습 - 신경망분석	퍼셉트론(Perceptron)	단층다층, 신경망분석 입은활을 향시롭게	○	<p>인간의 신경망과 유사하게 만든 입력층, 은닉층, 출력층으로 구성되어진 학습 능력을 가진 신경망 모델</p> <p>[유형]</p> <ul style="list-style-type: none"> -단층퍼셉트론: AND, OR, NAND -다층퍼셉트론: XOR(리거 가능, 은닉층추가, bias(편향)) <p>[기초요소]</p> <ul style="list-style-type: none"> -Bias Layer: 기초데이터 입력 -Input Layer >가중치, Bias(편향), Net Input Function(가중치 입력값 곱), 활성화 함수, Critical Point(활성화 최소값) -Output Layer: 출력값 <p>[활성화함수] 일정 범위값 범한 전달, 단조 증가 함수</p> <ul style="list-style-type: none"> -항상 합수: $f(x) = x$, 선형 출력 -시그모이드 합수: $y = 1/(1+e^{-x})$, 0과 1사이, 0.5기준 -ReLU함수: $ReLU(x) = 0 \text{ if } x < 0, 1 \text{ if } x \geq 0$, 0보다 크면 x -Gated 단단수: $f(\text{NET}) = 1 \text{ if } \text{NET} \geq T, 0 \text{ if } \text{NET} < T$, 기준값 T기준 0 아니면 1 	
6	지도학습 - 신경망분석	파라미터/하이퍼 파라미터	파기서걸 학습환경 매그栅 네	○	<p>[파라미터(Parameter)] 모델에 학습으로 결정</p> <ul style="list-style-type: none"> -파라미터는 최적화하는 모델 구현 -종류: 선형회귀(선형회귀분석), SVM의 SV, 선형회귀에서 결정계수 -주제: 편향 <p>[하이퍼파라미터(Hyperparameter)] 사용자, 허리스틱 기반 결정</p> <ul style="list-style-type: none"> -목적: 모델링 최적화 파라미터 도출 -종류: 학습률, 경사하강법 반복횟수, 활성화 함수 -조작 가능 <p>[하이퍼파라미터의 튜닝 방법]</p> <ul style="list-style-type: none"> -Manual Search -Grid Search -모든 조합 탐색, 시행횟수 한계(전수 탐색한계) -Random Search -랜덤워밍(방위내 우직위 반복), 탐색 범위 부여(최대최소) -Bayesian Optimization -관측데이터 기반 F(수정Bayesian 활용), Acquisition Function(학습 기반 입력 후보추출) 	
7	지도학습 - 신경망분석	활성화 함수	활시롭게 핫소	○	<p>[활성화함수] 일정 범위값 범한 전달, 단조 증가 함수</p> <ul style="list-style-type: none"> -항상 합수: $f(x) = x$, 선형 출력 -시그모이드 합수: $y = 1/(1+e^{-x})$, 0과 1사이, 0.5기준 -ReLU함수: $ReLU(x) = 0 \text{ if } x < 0, 1 \text{ if } x \geq 0$, 0보다 크면 x -Gated 단단수: $f(\text{NET}) = 1 \text{ if } \text{NET} \geq T, 0 \text{ if } \text{NET} < T$, 기준값 T기준 0 아니면 1 <p>[원인인코딩과 소스트레스스]</p> <ul style="list-style-type: none"> -원인인코딩 데이터를 수수께끼 데이터로 변형, 정밀 인코딩(One-hot Encoding) >파이썬의 유형에 따라서 세로은 고유값 추가, 해당하는 컬럼 1로 처리 >세상값 수치화, 수학적 흐른백터 >차원의 저주문제: 차원이 증가할수록 학습데이터 줄어들고 성능 떨어짐 -출처: 원인에서 정규화(0-1로 처리)처리 소프트웨스 >0~1사이의 실수, 확률값 -출처의 출처은 항상 1 	
8	지도학습 - 신경망분석	경사 하강법 (Gradient decent)	파베미	○	<p>함수의 기울기(경사)를 구하여 기울기가 낮은 쪽으로 계속 이동시켜 글 값에 이를 때까지 반복시키는 1차 근사값 발견을 위한 최적화 알고리즘</p> <p>[구조설명]</p> <ul style="list-style-type: none"> -기초 -Initial Weight: 초기가중치 -Gradient: 기울기(경사) -Global Cost Minimum: 최저값 -Stop Condition: 학습완료조건 <p>[알고리즘]</p> <ul style="list-style-type: none"> -특수한 경사 하강법 (Stochastic Gradient Descent): 하강법회귀함수 -매번 경사 계산, 데이터학습마다 간접, 지역최저점 빠질수 있음 -매번 경사하강법 (Batch Gradient Descent) -하나의 배치를 들어서 하강, 전체데이터 학습후 하강, 수렴이 안정적 -미니배치 경사하강법 (Mini-Batch Gradient Descent) -전체 배치 사이즈(시그모이드)로 나누고, 각 배치로 학습, SGD보다 적은 계산 <p>*Local Maximum/Minimum & Global Maximum/Minimum 문제</p> <ul style="list-style-type: none"> -local maximum 문제: 다른 최적화 문제로 해결해야 하는 경우에 발생하여 global minimum 찾음. -Gradient Descent: MLP(Multi Layer Perceptron) 학습시, 학습 layer가 깊어질수록 기울기가 소 면 기울기 초기화가 사라지며 Gradient Vanishing Problem 발생 	
9	지도학습 - 신경망분석	기울기소실 문제 (Vanishing Gradient Problem)	활가 학활가 기기애	○	<p>MLP에서 Hidden Layer의 학습 layer가 깊어질수록 기울기가 소멸, 학습 효과가 사라지는 현상</p> <p>[원인]</p> <ul style="list-style-type: none"> -MLP(Multi Layer Perceptron) <p>[원인]</p> <ul style="list-style-type: none"> -활성화함수(시그모이드함수): 소실(0또는 1에 수렴) / 폭주(임계값을 넘어간 기울기) -가중치초연(기중치영향): 역전파증 가중치폭주 / 모델에 적합하지 않는 가중치 -학습률 -학습률 초기화 >미분값보전 ReLU함수 활용 >음의 값 활용, Leaky ReLU사용 >은닉층에서 시그모이드 합수 사용 -가중치증연 >기울기자르기(Gradient Clipping): 임계값 넘지 않도록 -기울기초연: 모델의 적합한 가중치 적용 -배치규화: 입력을 평균과 분산으로 정규화 	

NO	포지	종포지	키워드, 두율	기본	서브노트	구성도 및 추가정보
10	지도학습 - 신경망분석	과적합(overfitting)문제	학류분대 복 측제드감분	O	<p>과적합(overfitting)문제: 감독학습(Supervised Learning)에서 과거의 학습데이터에 대해서는 잘 예측하지만 새로 들어온 데이터에 대해서 성능이 떨어져서 일반화가 어려운 문제</p> <p>[발생원인] -지나친 학습: Overfitting -지나친 투닝: Over-fitting due to Noise -분류 문제: 범주 별 데이터들을 잘 분류하지 못한 경우 -데이터 부족: Over-fitting due to Insufficient example -복잡한 모형: unnecessarily complex model</p> <p>[예방방안] -차원축소: PCA(Principal Component Analysis): 주성분분석, 3차원 > 2차원 -Regularization(제한): 맵다방에 추가 제널티 부여 출력층 사용 -DropOut: 득점 은닉층 제거 또는 데이터 0으로 만들어 누락 -감소: 모수(parameter)의 수 감소, 은닉층 감소 -분석/검증: 양자분류/Cross Validation</p>	Overfitting과 Underfitting
11	지도학습 - 신경망분석	드랍아웃	Back Propagation, Overfitting, 과적합(Overfitting), 비례화학적	O	<p>실증신뢰 및 학습을 위해 오류전파(Back Propagation) 수행 시, 과적합(Overfitting)문제를 해결하기 위해 은닉층의 일부 노드를 무작위로 비활성화 하는 기법</p> <p>[유형] 과적합 -feed forward: 은닉 초기 입력 노드 대상 -Back Propagation: 은닉 훈련 노드 대상 -시간공간: 드롭아웃 시간공간적 임관성 고려 -hybrid: Feed + Back, 입력+출력</p> <p>[작동원리] 일비화학적 -DropoutRate 입력: 파라미터 입력, 0.5입력시 50%날립 -노드 비활성화: 은닉 훈련 노드 선택해 학습기준 비활성 -신경망 학습: 비활성화 상태 학습 -오류전파: 비활성화 상태로 하여 학습수행 -테스트: 비활성화 벡터로 테스트</p> <p>[특징] -모든 노드 대량, 동조현상 최저(증복처리), 양상플호과 -과적합 방지 <p>[고려사항] -속도: 인자, 무작위 제거발생, Batch Normalization을 통한 드랍아웃 대체 -기준화 활용: 레이어전달 데이터 미흡, 가중치값을 드랍되지 않는 노드에 추가적용</p> </p>	
12	지도학습 - 분류(Classification)	CNN	CPFRDB	O	<p>데이터의 대비를 높여서 특징 추출 분석</p> <p>[동작방식] -Convolution Layer를 통한 득점(feature) 추출 -Pooling Layer를 통한 차원 축소 -Fully Connected Layer를 통한 최종 분류</p> <p>[Layer 구조] -Convolution Layer > 합성곱(Convolution) 연산을 통해 Feature를 추출하는 레이어 -Pooling Layer > Sub Sampling을 통한 차원의 축소 -Fully Connected Layer > Convolution, Pooling 처리결과 <p>[성능 개선] -ReLU: 0~1 확성화 함수, 기울기소실 Overfitting 방지 위해 득점 뉴런 미동적 학습 수행 -Dropout: 인공 신경망의 Overfitting 방지 위해 득점 뉴런 미동적 학습 수행 -BiGdata: 과적합(Overfitting) 문제 해결, 여러 데이터의 경험 누적</p> </p>	
13	지도학습 - 분류(Classification)	Pooling Layer	입영연이반결 평화흔적	O	<p>Convolution Layer에서 출력 데이터를 일정으로 밟아서 Sub-Sampling을 통해 출력 데이터 [Pooling]의 크기를 줄여서 특정 데이터를 강조하는 레이어</p> <p>[특징] -파라미터감소: 과적합 방지, 예산감소 -축소: 계산 수 유지, 득점 추출, 변화적용 <p>[동작] 1. Feature Map 입력: Convolution Layer의 출력 값인 Feature Map을 제날 개수 2. Pooling 영역 설정: 커널 사이즈 만큼 3. Pooling 연산: Pooling Average, Max 4. Pooling 영역 이동:Stride 값 만큼 window를 이동해 Pooling 영역을 재설정 5. 반복: 2 ~ 4번 반복해 세밀별 Activation 출력을 생성 6. 결과: 출력 = 세밀별로 생성된 Activation을 FC Layer의 입력으로 출력</p> <p>[유형] -Average: 평균, 득점이 약한, 결과가 안정음. -Max: 득점이 강하지만 overfitting 되기 쉬움 * 약점 보완하기 위해 두 기법 혼용 -Mix: 랜덤하고 평균/최대 반복 -Stochastic: 임의값 선택</p> </p>	
14	지도학습 - 분류(Classification)	RNN(Recurrent Neural Networks)	순환구조	O	<p>내부에 순환 구조를 가진 입력과 출력을 시퀀스 단위로 처리하고 BPTT(Backpropagation Through Time) 이용하는 인공 신경망 모델</p> <p>[특징] -뉴런 사이의 연결 Directed Cycle을 형성해 자신을 거리며 Recurrent Weight를 포함 -인공 신경망의 구조 -Cell: 이전 값과 기억을 유지하는 신경망 알고리즘 -은닉층: 처리후 출력보내고 다시 가중치 값으로 사용</p> <p>[학습원리] -Cell 저장: 이전값을 기억함. -Recurrent Weight: 반복가중치 구조 -BPTT: 오류역전파, 모든 시점에 대한 가중치 공유함, 기울기소실 LSTM으로 개선 -Sequential Data: 개별의 내용과 함께 문맥 이해, 길어질수록 전달 한계 발생</p> <p>[학습과 시각, BPTT(Backpropagation Through Time)] -기울기 역전파 방법에 시간이 추가되어 BPTT(BackPropagation Through Time)이라는 변형된 학습 방법을 사용 -다면, 역전파의 거리가 늘어나면 gradient 값이 죽어버거나 사라지는 현상이 발생하는 문제점이 발생(이제 역전파 구조가 LSTM과 GRU임)</p> <p>[구조] 입력 > 은닉 > 출력 -one to one - 기본 NN 구조 -one to many - 이미지 캡션 (이미지 -> 문장) -many to one - 감정분석 (문장 -> 흐름/분위) -many to many - 기계번역 (문장 -> 문장), 비디오 캡션 (비디오 -> 문장)</p>	
15	지도학습 - 분류(Classification)	LSTM(Long Short Term Memory)	정연 개설피모	O	<p>RNN의 Hidden Layer를 Input Gate, Output Gate, Forget Gate라는 세 가지 게이트를 이용하여 장기· kısa· 외온· 외온· 문제를 해결한 알고리즘</p> <p>[특징] -도입 초기 네트워크 레이어 -4개의 상호작용 가능한 GATE 구조 <p>[구조요소] 정보지장작제 -Gate -Cell State에 정보 주사작제 -Forget Gate: 합수를 통한 3개 Gate, 정보 통과 여부 결정, 셀 상태 보호 및 제어 -Cell State -이전 정보· 신령연산 ·> 다음단계 전달 -메모리작: 전자적 시술, 관통, 정보 전달 -다음 연산: 기울기 소신문제 해결</p> <p>연산처리 -Forget Gate: Cell State에서 어떤 정보를 제거할지 결정, 시그모이드 합수로 구성 -Input Gate: 새로운 정보를 cell state에 저장할지 결정, tanh 합수 사용 -Memory Update: 과거 시점의 셀 상태 경신 -Output Gate: 최종 출력값을 결정, cell state를 tanh 합수를 거쳐 1과 1사이 값을 생성</p> <p>[GRU] Gated Recurrent Unit -LSTM과 같은 구조 -Gate 2가(update, reset), 빠름</p> </p>	
16	지도학습 - 분류(Classification)	서포트 벡터 머신 SVM(Support Vector Machine)	하시미스 커트하소	O	<p>학습데이터를 두 개의 클래스로 나누는데 Margin을 최대로 하는 결정직선을 찾는 분류(classifier) 알고리즘</p> <p>[구조요소] 하시미스 -초평면(Hyperplane): 클래스를 구분하는 면/선/곡선), 분류, 회귀분석 -Support Vector: 학습 데이터 중에서 분류 경계에 가장 가까운 점에 위치한 데이터 -Margin: 학습 데이터 중에서 분류 경계에 가장 가까운 데이터로부터 분류 경계까지의 거리 -Slack Variables: 여유변수, 선형분류를 위해 허용하는 오차변수, Soft Margin SVM 사용</p> <p>[구조요소] 커트하소 -기울기와 y截距를 3차원 공간으로 변환하여 경계면 찾기(2차원>3차원) -기울기 턱: 커널 합수 사용시 연산부하를 해결하는 기법 -하드 마진 SVM(Hard Margin SVM): 오류를 허용, 하드마진 적용 어려움으로 주로 이용 -소프트 마진 SVM(Soft Margin SVM): 오류를 허용, 하드마진 적용 어려움으로 주로 이용 *Support Vector Machine은 선형으로 분리가 가능하지 물가능한지에 따라 적용하는 방식이다 <p>[학습 분석] -하나니 그 이상의 독립변수의 종속변수에 대한 영향의 추정을 할 수 있는 통계기법 -regression analysis</p> </p>	

NO	포지션	종류	키워드, 두율	기본	서브노트	구성도 및 추가정보
17	지도학습 - 분류 (Classification)	KNN (K-Nearest Neighbor)	다수계단 평가리다 유마코	O	<p>Sample에 주어진 x에서 가장 가까운 k개의 원소가 많이 속하는 class로 x를 분류하는 비모수적 학습법으로 추정법</p> <p>[특징] 다수계단법</p> <ul style="list-style-type: none"> -최고인접 다수점: 기준 데이터 중 가장 유사한 k개의 데이터를 측정하여 분류 -유사도(거리)기반: 유사도를 입력 값이 들어온 후 유사도를 계산, 모델 만들지 않음. -단점 유연성: 모형 단순하여 파라미터의 개수가 거의 없음 -NN(Nearest Neighbors) 개선 >KNN은 가장 근접한 k개의 데이터에 대한 다수결 내지 가중합 방식으로 분류 >기존 유사도는 유사도를 계산해서 그와 같은 class에 일방적으로 분류, 합집 데이터 성능 구름 <p>[축약식] 평가리다</p> <ul style="list-style-type: none"> 1. 사용자로 fingerprin 입력(음표, 동그라미) 확인 2. 거리(반경) k개 데이터를 training set에서 추출 3. 추출 데이터들로 클래스 및 label 확인 4.(다수결(Majority Voting)에 의한 클래스 예측 <p>[기준기반 알고리즘] 유마코</p> <ul style="list-style-type: none"> -유사리다(안경)과 점간 거리를 계산하는 문제에서 두 집단 사이의 거리 -마지막으로 각각의 거리를 계산한 다음 두 집단을 판별하는 문제에서 두 집단 사이의 거리 <p>[장단점] 단점하고 흐름, 성능제한 점은?</p> <ul style="list-style-type: none"> -정상: 유사도(거리)를 계산하는 데 시간이 걸리고 계산량이 많아서 그와 같은 class에 약간 -단점: 성능(기반점/높은 차원으로 구현)과 비용(시간/장비/부정확) 	
18	비지도학습	데이터 레이블링 (Data Labeling)	네이프아이 생분류관		<p>기계학습 분야에서, 머신러닝이나 딥러닝 모델링 학습 전에 학습 데이터에 특정 값을 부여 해주는 행위</p> <p>[방법] 내부화하고</p> <ul style="list-style-type: none"> -내부화: 라벨링: 내부 자체 라벨링, 대규모 조직 적합, 리소스부족 -합성화: 레이블링: 레이블링 프로젝트 생성, 비용이슈 -프로그램: 방식의 라벨링: 자동화 처리, QA 기술이슈 -아웃소싱: 라벨 전문 그룹, 위탁, 고용한 라벨, 비용이슈 <p>[제작] 자동화 생분류관</p> <ul style="list-style-type: none"> -자동화: 데이터 세트를 비롯한 출처, 블로그, 딥러닝, 레이블링 프로그램 -고객: 레이블링 데이터 분류, 딥러닝, 레이블링 프로그램 -제작: 데이터 적용 및 평가: 인과관계 분석, 응용 도메인을 통한 결과 검증 -유지 및 관리: 데이터 적용, 관리 및 시각화, 품질점검 <p>[활용]</p> <ul style="list-style-type: none"> -음성: 음성인식 라벨링(대화형 응답 받아 적기) -이미지: 시각인식 라벨링(마이크, CCTV) -텍스트: 텍스트 라벨링(서치) 	
19	비지도학습	DBSCAN (밀도 기반 클러스터링)	에코보노 에인코보	O	<p>밀의 풀리면서 중심을 이동시키며 중심으로부터 정해진 반경 거리 내에 최소 데이터 포인트 개수를 확장하여 밀도 기반으로 군집화 수행하는 알고리즘</p> <p>[특징]</p> <ul style="list-style-type: none"> -밀도 학습: K-Means와 같이 클러스터(군집) 수를 미리 정하지 않아도 됨 -밀도 연결: 밀도에 따라 클러스터를 연결, 기하학적 모양을 가지는 군집도 찾을 수 있음 -Noise Point: Noise Point를 통해 Outlier 검출 가능 <p>[구조요소] 에코보노</p> <ul style="list-style-type: none"> -Epsilon: 중심으로부터 거리 -Core Point: e(Epsilon)내에 데이터가 m개 이상 존재하는 포인트, 군집의 중심 -Border Point: 군집에 속하는 데이터 -Connected: core point와 core point가 반경에 걸친 경우 -Noise Point: 어떤 점의 중심으로도 조건을 만족시키지 못하는 데이터 <p>[활용식]</p> <ul style="list-style-type: none"> ① Epsilon 설정: 두 인스턴스 최대 허용 거리 ② minPts 설정: Epsilon내 인스턴스의 최소 개수 ③ Core point 분류 ④ Border Point 분류 	
20	비지도학습	K-Means	거리기반, KCentroid	O	<p>n개의 데이터를 K개의 군집으로 분류하기 위해 거리 기반으로 계산해 나가는 Clustering 알고리즘</p> <p>[구조요소]</p> <ul style="list-style-type: none"> -K값: 클러스터링하여 묶을 클러스터의 개수 -Centroid: 클러스터의 중심 <p>[장단점]</p> <ul style="list-style-type: none"> -단점: 배수, 대용량 처리, 단양성 -기준: 정의 어려움, 초기클러스터링 어려움, 결과난해 	
21	비지도학습	자기조직화지도 (Self-Organizing Map)	임기경노를 거친수반		<p>차원 축소(dimensionality reduction)와 군집화(clustering)를 동시에 수행하는 디멘시널리티 감마피지의 학습 과정을 모델화한 인공신경망 알고리즘</p> <p>[특징]</p> <ul style="list-style-type: none"> -차원 축소: Dimensionality Reduction -군집화: Clustering -승수: 반복을 통해서 여러 가지 카테고리와 유사한 카테고리를 찾는다. -빠른속도: 전방으로 전달해서 즉시 분석, 실시간 학습 가능 <p>[구조요소] 임기경노</p> <ul style="list-style-type: none"> -입력층: 입력, 차원을, 경계를, 단양성 -경계: 차원에서 배운, 대용량 처리, 단양성 -기준: 정의 어려움, 초기클러스터링 어려움, 결과난해 	
22	비지도학습 - 잠재변수 모델	PCA(주성분분석) (Principal Component Analysis)	전공과부 공행복사 페카풀 다		<p>고차원 데이터를 저차원의 데이터로 환원시키는 기본 선형 관계에 있는 저차원공간(주성분)의 표본으로 변환</p> <p>[구조요소]</p> <ul style="list-style-type: none"> 1. 데이터 선형화: 데이터 정규화 수행 2. 공분산 행렬 계산: 공분산 행렬 계산 3. 고유값, 고유벡터 계산: 공분산 행렬 분해를 통한 고유값, 고유벡터 계산 4. 데이터 두동: 저차원 데이터 두동 <p>[기준요소] 공행복사</p> <ul style="list-style-type: none"> -공분산: x_i의 흩어진 평균 얼마나 서로 상관관계를 가지고 흩어졌는지를 나타냄. ->로 상관관계가 있을 경우 공분산은 0, -공분산 행렬의 데이터 좌표평면 상의 공분산 값은 원으로 하는 행렬 -eigen Value: 공분산 행렬 A를 선형 변환한 결과가 차기 Eigen Value로 표기되는 행렬 -eigen Vector: 공분산 행렬 A를 선형 변환한 결과가 차기 Eigen Vector로 표기되는 행렬 <p>[활용식] 페카풀다 PICA FDLA</p> <ul style="list-style-type: none"> -PCA(Principal Component Analysis): 주성분분석, 차원削減, 주성분 분석 -ICA(Independent Component Analysis): 차원 증가, 독립성에서 서로 다른 특징 추출 -FDA(Fisher Discriminant Analysis): 구별적인 특징 추출 및 분류학 -LDA(Linear Discriminant Analysis): 데이터를 축소해서 최적화 	
23	비지도학습 - 잠재변수 모델	GAN (Generative Adversarial Network)	제디파트 편진봉	O	<p>생성자(Generator)와 판별자(Discriminator)가 서로 경쟁(적대적)하는 과정을 통해 정보를 학습하는 대표적 비지도학습</p> <p>[구조요소]</p> <ul style="list-style-type: none"> -적대적 모델: 2개 모델을 통한, 새로운 학습을 만드는 프레임워크 -Min-Max 문제: 한쪽은 유틸리티이고, 반쪽은 구분하는데 확률을 높이는 과정 <p>[기준요소]</p> <ul style="list-style-type: none"> -Generator: Zero-Mean Gaussian 노이즈 기반 Fake Sample 생성 -Discriminator: Real Sample와 Fake Sample 구분, 각 확률 Estimate -Fine Tune Training: Generator, Discriminator 학습 오차 보정 -Training Data: 학습데이터 <p>[GAE(내부적)]</p> <ul style="list-style-type: none"> -생성자: Generator와 Discriminator 성능차이, DCGAN 활용 -모든 전동: 속도 속이는 학습만 진행, 학습 기억 사용 -모든 봉지: Generator가 Discriminator를 속이는것만 집중 	
24	비지도학습 - 잠재변수 모델	DCGAN (Deep Convolutional GAN)	컨베풀풀 RLT	O	<p>GAN의 Fully Connected Layer를 제거하고 Convolution Layer와 배치정규화 구조를 사용하여 안정적인 학습이 가능한 알고리즘</p> <p>-Convolution 적용, Pooling 적용 X</p> <p>[기준요소]</p> <ul style="list-style-type: none"> -네트워크 구조 -2D 이미지 -배치정규화: 활성화 함수 간의 정규분포화(평균 0, 분산 1), 기울기감소 해결 -Fully-connected hidden layers 삭제 -Pooling Layer 사용안함 -활성화 함수 -ReLU(Rectified Linear Unit): 생성자층 사용 -Tanh: 활성화 사용 -Leaky ReLU: 판별자층 사용, 음수 ReLU 	
25	강화학습	Q-러닝	Model Free, Q- 정밀도 최적화 반복 테크 최적화법	O	<p>Q-Learning은 Model이 없어(Model-Free) 학습하는 강화학습 알고리즘</p> <p>[구조요소]</p> <ul style="list-style-type: none"> -강화학습, Model-Free -현재 상태에서 향후 행동의 보상에 대한 미래의 보상들의 종합적인 품질(Quality) 기대값. <p>[기준요소]</p> <ul style="list-style-type: none"> -정책: 최대보상(최고 Q값에 따른 액션), 미래보상(현재보상 기대) -밸런스정점식: 정책반복(최적 보상 찾기), 재귀함수(현재최고보상, 미래보상) -Q러닝: 태이프 기반(법칙적 반복), 반복적 근사(반복을 통한 Q근사) -밸런스 방법식은 현재 상태의 가치 함수와 다음 상태의 가치 함수 사이의 관계 <p>[활용]</p> <ul style="list-style-type: none"> ① value table Q 초기화 ② 정책 기반 Action 선택, 신호수송, 관찰 ③ 새로운 상태로 최대보상 업데이트 ④ 새로운 상태로 최대보상 설정, 반복수행 <p>[강화학습]</p> <ul style="list-style-type: none"> -Deep Q-Learning: 기존 Q-러닝에서 Q테이블을 딥러닝 CNN으로 변경한 것임. -Markov Chain: 마르코프 청정(특정 미래 상태의 확률은 오직 과거의 상태에 의존) 	

NO	포지	종포지	키워드, 두음	기본	서브노트	구성 및 추가정보															
26	모델평가	혼동 행렬 (Confusion Matrix)	혼동 행렬	O	잘못된 예측, 분석을 파악하기 위해 예측 값과 실제값의 일치 여부를 행렬로 분류하는 모델 평가 기법	<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">실제 정답</th> </tr> <tr> <th colspan="2"></th> <th>Positive</th> <th>Negative</th> </tr> </thead> <tbody> <tr> <th rowspan="2">Positive</th> <th>Positive</th> <td>True Positive</td> <td>False Positive (Type 1 Error)</td> </tr> <tr> <th>Negative</th> <td>False Negative (Type 2 Error)</td> <td>True Negative</td> </tr> </tbody> </table>			실제 정답				Positive	Negative	Positive	Positive	True Positive	False Positive (Type 1 Error)	Negative	False Negative (Type 2 Error)	True Negative
		실제 정답																			
		Positive	Negative																		
Positive	Positive	True Positive	False Positive (Type 1 Error)																		
	Negative	False Negative (Type 2 Error)	True Negative																		
【혼동행렬】 -정답 True Positive(TP): 실제 값이 Positive이고, 예측 값도 Positive인 경우 True Negative(TN): 실제 값이 Negative이고, 예측 값도 Negative인 경우 -오답 False Positive(FP): (1종 오류): 실제 값은 Negative이나, 예측 값이 Positive인 경우 False Negative (FN): (2종 오류): 실제 값은 Positive이나, 예측 값이 Negative인 경우																					
27	모델평가	정밀도(Precision)/재현율(Recall)	정체정	O	[정밀도-Precision] TP/(TP+FP) -모델이 맞는 것을 예측한 경우에서 실제 True인 것의 비율 -정답이라고 예상했을 때, 실제 정답인 것의 비율 ex) 정답 문서에서 정답으로 예상한 문서들의 비율 [재현율-Recall] TP/(TP+FN) -실제 True인 것 중에서 모델이 True라고 예측한 것의 비율 -실제 정답인 것들에 대해 예측한 경우(TP), 정답을 놓친 경우(FN)의 비율 -진짜 정답 중 모델이 정답이라고 예측한 것들의 비율 ex) 정답 문서에서 재현율(recall)은 관련 있는 문서들 중 실제로 검색된 문서들의 비율 [정밀도 Accuracy] Baseline TP/TN / (전체 문서 수)																
28	파괴적망각	파괴적망각 (Catastrophic Forgetting)	정증동	S	Single Task(단일 파괴) 성능우수, 다른 종류의 Task를 학습하면 이전에 학습했던 Task에 대한 성능이 현저하게 떨어지는 문제(이전학습망각)																
29	기계독해	기계독해 (Machine Reading Comprehension)	질매유답 검독 지생후	S	주어진 질문에서 인공지능(AI) 알고리즘 스스로 문제를 분석하고 질문에 최적화된 답안을 찾는 기술 [기계독해 과정] -사용자질문>문서에 대한 매트릭스구성>유사도점수부여>답변생성 +검색 단계 -TF-IDF: 어떤 단어가 문서내에서 얼마나 중요하지 여부 >Term Frequency-Inverse Document Frequency -TF-IDF 매트릭스구성 -BM25: 주어진 문서와 문서연관성 측정 합수 알고리즘 -Point Network: 문장 내의 다른 단어를 보고 힌트를 얻어 현재단어에 대한 정보 index 출력 네트워크 [기계독해 유형] EDM -Exact Match Answer Datasets: 질문에 답변이 존재 -Description-Negative Answer Datasets: 답변을 생성 -Multiple-Choice Datasets: 답변 후보군 중 선택 [기계독해평가] 정밀도 재현율 F1 -Exact Match: 정답을 찾을 때마다 점수를 더함 -F1 Score: 2*원래F1 / -ROUGE-L: 재현율 / -BLEU: 정밀도 -HQE(Human Equivalence Score): 대화형, 신뢰있는 답변																
30	ELECTRA	ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately)	생성적 GAN, 제디제디	S	Generator 가 문장에서 미스킹된 위치에 올바른 토큰을 예측하고 Discriminator 는 입력 받은 문장이 Generator에 의해 대체되었는지를 맞추며 학습하는 모델 -Generator: 데이터 생성, Replace Token 수행 -Discriminator: 모든 토큰 판별 -Generator Loss: 단일 학습으로 구분할 수 없는 간접 손실 -Discriminator Loss: 만들면서 성립도 실험과 더러운 딥으로 처리 +GAN 과정: 차이점은 Generator가 적절한 학습을 하지 않고 생성적으로 학습 +생성 모양이 원본 토큰을 생성하는데 상공, 그 토큰은 'fake'가 아닌 'real'로 간주 +시모모이드만 사용(비율, 0이나 1이나, 몇나 몰려나)																
31	AI유형	연합학습	중모학전글	O	분산된 환경에서 디비터의 적극적인 공유 없이 많은 토클선서버들이 중앙서버의 지시를 받아 혼란을 수습하는 학습 [중모학습] -정답보조: 분산서버가 디비터를 내보내지 않음 -제작연필로운집-결과로 모델만 전송 [분석] -중간 데이터는 데이터 자체로 학습 -수집: 데이터 A 데이터 B 영역, 수직으로 충돌 -수령: 데이터 A 데이터 B 영역, 수평으로 교차 -선언: 교차한 영역 리밸 전이 [학습환경] -Cross Silo: 지역적으로 벌어진 센터, 1~1000 -Cross Device: 수 많은 모바일/iot기기, 10~1000상 *종종은 하시온 네이버 같은 군사 불가																
32	AI유형	GNN	AUR	S	그래프 신경망, 노드와 엣지로 구성된 그래프 형태 학습 [구성요소] -노드: 노드 분류, 노드 회귀 -엣지: 연결 여부, 연결 관계 -그리프: 네트워크 구조, 그리프회귀 [학습단계] -Aggregate: 이웃 노드들의 은닉변수 정보 취합 -Update: 타겟 노드와 연결 결합된 노드들의 Update -Readout: 모든 노드 은닉변수 결합하여 그래프단위 은닉변수 생성																
33	AI공격	최대적공격	오퍼전추	S	-온도공격(Poisoning) >학습데이터 악의적인 데이터 주입 >회피 공격과는 다른 학습과정 공격 -회피공격(Evasion) >원본 데이터에 perturbation(작은변화) 추가 >모든 학습 데이터를 피해가며 학습 -원본공격(Imperceptible) >모든 데이터를 깨끗, 덜어서 학습데이터 유출 복원 -추출공격(Extraction) >모델에 희미한 흔적을 던져 모델 자체를 알아남, 모델복사																
34	학습용데이터	AI학습용데이터 비용산정가이드	비라구 설공인제적최	S	[비설정 3초] -비용산정요소: 비용발생 항목, 물량, 인원, 기간, 장소, 장비 -작성 그룹요소: 최소사업단위, 단가 -데이터구축공정: 구축유형(텍스트, 이미지, 음성, 비디오) 구축 프로세스 [비설정필자] 설공인제적최 -작업 설계: 작업요소설계, 물류산정 -주제설계: 주제선택, 주제설정 -작업구성: 투입수기준비, 비용산정, 수집, 생산, 가공 -제작비산정: 적절인력비기준, 제작비산정, 15~20% -직장경비산정: 직장인력비기준, 콘텐츠구매, 장비, 장소 -최종비용산정: 직장인력비 + 제작비 + 적립경비																
35	학습용데이터	메타학습	거모희, 쥬넷	O	이것과 모르는것을 족감으로 구분하는 메타인지로부터 시작한 개념, 적은 양의 데이터로도 학습하고 문제에 적용하는 학습방식 [유형] -기저기반학습방식: Prototypical Network, Relational Network -모델기반학습방식: Memory이용, RNN -최적화학습방식: Model Parameter 최적화 *학습률적응																
36	학습용데이터	튜셔닝(Few-Shot)	제원류	S	소량의 데이터로 효과적인 학습을 하는 학습 방식 [유형] -세 가지 튜셔닝: 한번도接触하지 않은 클래스 학습, 호랑이>사자>라이언 -일반 튜셔닝: 하나의 샘플만 모아두고, 한개의 호랑이, 두리째는 맞춤 학습방법, 풀샘플과 맞춤 학습방법 [구성요소] -Input: Word2Vec / Word Embedding -Encoder/Decoder Block >Self-Attention: 각 단어의 vector간 연관도 점수화, Query, Key, Value 벡터 생성 >Multi-Head Attention: 가치층을 투여 다른 조건으로 학습, 더 많은 후보군 제공 >Add&Normalization: 추가 및 정규화 >Encoder/Decoder Attention >Output: Linear > Softmax >계산단위의 softmax를 적용해 합이 1이 되도록 한다																
37	트랜스포머	트랜스포머	입백임 인디셀링합 출리소	O	리볼루션 필요는 빠른 AI, CNN, RNN 대체 > 브랜스포머 -자연어는 알림, 명령, 명령, 연사처리 -자연어는 의미를 찾는 결코 아름다운 [구성요소] -Input: Word2Vec / Word Embedding -Encoder/Decoder Block >Self-Attention: 각 단어의 vector간 연관도 점수화, Query, Key, Value 벡터 생성 >Multi-Head Attention: 가치층을 투여 다른 조건으로 학습, 더 많은 후보군 제공 >Add&Normalization: 추가 및 정규화 >Encoder/Decoder Attention >Output: Linear > Softmax >계산단위의 softmax를 적용해 합이 1이 되도록 한다 * NLU의 학습 -각 단어들의 query와 key vector 연산을 통해 관계를 유추하기에 단어의 거리는 무관 -행렬로 병렬로 인식하므로 빠름																

NO	로직	종로직	키워드, 두음	기본	서브노트	구성도 및 추가정보
38	Chat GPT	Chat GPT	김보희	○	<p>GPT-3.5 기반 대화에 최적화된 대화전문 인공지능 -강화학습, 대화학습속성</p> <p>[질자] 김보희 -Step1 보상정책학습: 학습데이터 샘플링, 행위 라벨링, Fine Tune GPT 3.5 -Step2 보상정책학습: 보상학습 데이터 샘플링, 순위부여, 보상모델학습 -Step3 강화학습(최적화): 신규데이터 입력, 결과생성, 보상평가, PPO수정정책데이트</p> <p>[기술요소] -RLHF(Rainforcement Learning with Human Feedback) >Agent(Action), Environment(Observation), Reward(오기서 유인 피드백) -PPO(Proximal Policy Optimization): 모델없는 강화학습</p> <p>[문제점] 현행상, 상식결여, 패턴의존, 제한적, 개인화부족</p>	<pre> graph LR A[대모데이터 수집 및 강화정책 학습] --> B[비교데이터 수집 및 보상 정책 학습] B --> C[강화학습 사용 보상정책 최적화] C --> D[신규 데이터 입력] D --> E[보상을 통한 결과 동양화] E --> F[보상 모델 학습] F --> G[보상 평가] G --> H[PPO 사용 정책 업데이트] H --> I[Fine-tune GPT-3.5] I --> J[행위 라벨링] J --> K[순위 부여] K --> L[보상 모델 학습] L --> M[보상 평가] M --> N[보상을 통한 결과 동양화] N --> O[신규 데이터 입력] O --> P[비교데이터 수집 및 보상 정책 학습] P --> Q[강화학습 사용 보상정책 최적화] </pre>
39	AI윤리	AI윤리	총금합, 인프라침 공연대책 인수	○	<p>[기획원칙] 인간 존엄성 원칙, 사회의 공공선 원칙, 기술의 합법적성 원칙</p> <p>[10대 윤리요건] 인권보장, 프라이버시 보호, 다양성 존중, 친해·금지 공공성, 대안성, 데이터 관리, 책임성, 안전성, 투명성</p>	
40	딥뷰	딥뷰	영내실내 수시대 예내예지		<p>사람처럼 인식하는 시각 인공지능 기술 -시각 인공지능 기술을 이용하여 다양한 사회 문제를 해결하기 위해 딥뷰를 사용</p> <p>[특징] 영내예대 -영어인식(사물과 사람분리), 내용파악(영상속내용파악), 실시간분석, 대용량분석</p> <p>[기술요소] 수시대 예내예지 +수시대 -시각장치: 시각데이터자산화, 빅데이터저장, API지원 -대규모처리: 처리파이프라인, 하이브리드스케줄러(GPU) +예지기반 -내용분석: 이미지분석, 시각텍스트노미, 시멘틱추론 -예지 활용: 예측기술, 실시간대출 -지향: 플랫폼</p>	
41	자연어	워드 일베팅	통합학 신워넷		<p>전체 단어들 간의 관계에 맞춰 해당 단어의 특성을 갖는 벡터로 바꿔주므로 단어들 사이의 유사도를 계산하는 기법</p> <p>[기술의 발전 측면 기술유형] 통합학 신워넷 -통계적 기반: TDM(Term-Document Matrix), TF-IDF, One-hot Encoding -신경망 기반: Word2Vec, BERT</p> <p>[현장 수용 기술 유형] -단어 벡터의 임베딩: Word2Vec, FastText, ELMo(Embedding from Language Models) -문장 수준의 임베딩: BERT, GPT</p>	
42	인공지능 평가	ROC Curve Precision-Recall Curve	알록민 경제정		<p>[ROC Curve] 알록민 모든 분류기에서 민감도(Sensitivity)와 특이도(Specificity)의 매개 변수를 표시함- 통해 분류 모델의 성능을 보여주는 그래프 +구성 -X축: 특이도(Specificity): FP / (FP+TN) -Y축: 민감도(Sensitivity): TP / (TP+FN) +AUC(Area Under the Curve) -ROC Curve 아래의 면적으로 환산하여 평가하는 방법</p> <p>[Precision-Recall Curve] 경계정 Y축: 정밀도(Precision)이고, X축이 재현율(Recall)을 기준으로 불균형한 데이터 셋일 때 성능 평가에 유리한 그래프 +구성 -X축: 정밀도(Precision) : TP / (TP + FP) -Y축: 재현율(Recall) : TP / (TP + FN)</p>	
43	인공지능 문제	편향(biased)	인습대 통고		<p>[유형] -Data 관점: 인간의 편향(Human Bias), 숨겨진 편향(Hidden Bias), 데이터 표본 편향(Data Sampling Bias) -Process 관점: 롱테일 편향(Long-tail Bias), 고의적 편향(Intentional Bias)</p> <p>[10가지] -연장화 데이터 학습 -복조 데이터의 디지털 가공 과정 -언어모델 모델 자체</p> <p>[대처] -정부 출판: 설명을 요구할 권리 규정, 자동화된 의사 결정을 제한할 권리 -기업 출판: 기업 AI윤리 수립, 정보 투명 전파 감리, XAI 개발 및 설계, biased 인식 알고리즘 개발</p>	
44	인공지능 문제	인공지능 보안	무기 오회전주		<p>[공격기법] 무기 오회전주 -무기성 출연: Poisoning attack(중독 공격, 오염 공격), Evasion attack(회피 공격) -기밀성 출연: Inversion attack(전도 공격, 학습 데이터 추출 공격), Model extraction attack(모델 추출 공격)</p> <p>[유익한 방지] -구조적: 적의적 공격 방지기법 >Gong: Defence-GAN >Poisoning Attack: 적대적 훈련 (Adversarial Training), 입력값 제한 >Evasion Attack: 이진 분류기 편별 >기밀성 허락점 공격 방지기법 >공종: Feature Squeezing >Model Inversion Attack: Gradient Masking >Model Extraction Attack: 퀴리 횟수 제한, Distillation</p> <p>[관련사례] -무기성 이용: 존행 오인식, 표지판 변형, 보험 사기, 임상 실험 회피 -공종 이용: 모션 퀴리, 일반화 결증</p>	
45	추천시스템	추천시스템	컨월립		<p>특정 시장에서 특정 사용자가 관심을 가질만한 리스트를 찾아주는 시스템 -특정 사용자의 관심사와 의사결정시스템에서 결선</p> <p>[추천 알고리즘] 컨월립 -컨텐츠기반: 사용기록기반, 첫사용자 불가 -유저필터링: 타사용자 사용기록 기반 -딥러닝용량: 딥러닝 기반 연구중</p> <p>[활기치료] -비즈니스: 매출, CTR(Click Through Rate) -품질지표: 연관성, 다양성, 새로운, 물에속성, 커버리지</p>	