

[0]요약

- 패턴 인식 및 회귀 문제를 풀기 위해 쓰이는 기계학습에 대한 전반적인 이론과 설계 방법
 - 기계학습
 - ◆ 신경회로망
 - ◆ 기저벡터머신
 - ◆ 모델 구축 시 고려할 문제점
 - ◆ 특징 추출 과정 방법, 성능 영향도
 - ◆ 딥러닝
 - 자가인코더, 제한볼츠만기계, 컨볼루션신경회로망, 회귀신경회로망
 - 딥러닝의 특징, 장점

[1] 서론

- 기계학습 일반적 순서(예)
 - 학습데이터(이미 종류를 알고 있는 과일)로부터 특징(feature: 색, 크기)을 추출하고 학습(training/learning)을 통해 클래스(오렌지, 자몽)를 구분하는 모델(결정 경계, 그래프의 점선)을 찾은 뒤, 새로운 데이터(새로 들어온 과일)의 특징 값을 기반으로 그 데이터의 클래스를 결정(test/generalization)한다.
 - 잘못 분류된 데이터(결정경계)를 완벽하게 구분할 수 있도록 하는데에서 발생하는 overfitting은 성능을 저하시킴.
- 해결할 수 있는 문제
 - 패턴인식(주어진 데이터 클래스 구분)
 - 회귀(연속적인 어떤 값 추정) or 함수 근사화
 - 학습 방법
 - ◆ 지도학습(패턴인식, 회귀 문제에서 학습 데이터의 클래스나 출력 값을 알고 이에 대한 피드백을 통해 학습)
 - ◆ 비지도학습(학습 데이터의 클래스나 출력 값을 이용하지 않고 데이터 패턴이나 클러스터, 밀도 등을 추정)
 - ◆ 반지도학습(클래스나 출력 값을 아는 데이터(labeled data)와 모르는 데이터(unlabeled data)를 함께 사용)
 - ◆ 강화학습(결과에 대한 피드백만 주어지고 정확한 클래스나 출력 값은 주어지지 않음)

[2] 기계학습 모델

- [2.1] 신경회로망(neural network)
 - McCulloch-Pitts neuron(1943)
 - ◆ 가중치가 곱해진 입력 값들의 합을 계산하여 그 합이 임계값을 넘으면 1, 아니면 0을 출력하는 인공 뉴런
 - Rosenblatt(1957)
 - ◆ 퍼셉트론(입력 값에 곱해지는 가중치를 학습하는 인공신경망 모델)
 - XOR과 같이 선형으로 분리되지 않는 문제는 해결 불가
 - 해결 : 다층 신경회로망(multi-layer neural network)
 - 다층 신경회로망(multi-layer neural network)
 - ◆ Hidden layer, hidden neuron 개수, activation function 선택 결정 필요함
 - Universal approximation theorem
 - 국소적으로 한계가 있고(locally bounded) 구간별로 연속인(piecewise continuous) 비다항식의 활성화함수를 가지는 한 층의 은닉뉴런으로 어떤 연속적인 함수도 학습할 수 있다.
 - 성능 측정
 - ◆ 평균제곱오차(mean square error)
 - ◆ 크로스엔트로피(cross-entropy)
 - 성능 학습
 - ◆ 기울기 하강(gradient descent)을 이용한 오차역전파법(error back-propagation)
 - ◆ 학습을 값이 클수록 학습 속도 빨라짐, 극값 수렴 못함, 발산 가능성 높아짐

- ◆ 국부최적화(local optimization) 문제
 - Global optimum에 도달하지 못하고 local optimum에 수렴
- ◆ 학습 데이터 입력 방식
 - Batch, mini-batch, online 등
- ◆ 학습 속도 향상
 - Error back-propagation 외 conjugate gradient backpropagation, quasi-Newton, Levenberg-Marquardt algorithm 등 존재.
- [2.2] 기저벡터머신(support vector machine)
 - ◆ 결정경계와 각 클래스의 데이터 간의 최소 거리로 정의되는 마진을 최대화한다는 목적
- [2.3] 확률 밀도 분포 추정법
 - ◆ 각 클래스의 확률 밀도 분포 추정
 - ◆ Parzen's window, Gaussian mixture model, hidden Markov model 등
 - ◆ 모수(parametric) 추정법
 - 최대우도추정(maximum likelihood estimation), 최대사후확률추정(maximum a posterior estimation), EM알고리즘(expectation-maximization algorithm) 등
 - ◆ 비모수 추정법
 - K-NN(k-nearest neighbor), 의사결정트리(decision tree)
- [2.4] 기계학습 모델 구축
 - 여러 선택지에서 같은 결과를 얻는다면 가장 단순한 방법이 가장 최적의 방법
 - 편향-분산 트레이드 오프(bias-variance trade-off) 문제
 - ◆ 평균제곱오차는 편향(bias), 분산(variance)로 나눌 수 있음(trade-off)

- 편향(bias)
 - 모델을 학습하는데 있어 학습 데이터를 얼마나 유연하게 받아들일 것인가에 대한 지표
 - 편향값이 높으면 기계학습 모델이 제대로 학습되지 않는 과소학습(underfitting) 문제 발생
- 분산(variance)
 - 학습데이터에 대한 모델의 민감도
 - 분산값이 높으면 학습 데이터에 포함된 노이즈까지 기계학습 모델이 학습했음을 의미 과도학습(overfitting) 문제 발생

[3] 특징 추출

- 주어진 문제를 풀기에 적합한 특징 추출(데이터 가공, 낮은 차원으로 변환)
- 데이터 기반 특징 추출
 - PCA(principal component analysis)
 - ◆ 주어진 데이터의 정보 손실을 최소화 하는 방향으로 특징 추출
 - LDA(linear discriminant analysis)
 - ◆ 주어진 데이터의 클래스 간의 거리를 최대화 하는 방향으로 특징 추출
- 데이터 고려하지 않고 특징 추출
 - 데이터를 다른 차원으로 변환한 뒤 변환계수(transform coefficient)를 취함
 - ◆ Discrete Fourier transform, discrete cosine transform, discrete wavelet transform 등
- 특정 도메인에 대한 사전 지식 기반 특징 추출
 - 음성 신호 MFCC(Mel-frequency cepstral coefficient)
 - 이미지 HoG(histogram of oriented gradient)