

4차 일석이조 조별보고서	
작성일: 2023년 9월 26일	작성자: 김동규
조 모임 일시: 2023년 9월 26일 8교시	모임장소: 구글미트
참석자: 이준용, 유정훈, 김동규, 이학빈, 탁성재	조원: 이준용, 유정훈, 김동규, 이학빈, 탁성재
구 분	내 용
<p>학습 범위와 내용</p> <p>(조별 모임 전에 조장이 공지)</p>	<ol style="list-style-type: none"> 1. 확률 기초 2. 베이즈 정리와 기계 학습 3. 최대 우도 4. 평균과 분산 5. 유용한 확률분포 6. 정보이론
<p>논의 내용</p> <p>(모임 전 공지된 개별 학습 범위에서 이해된 것과 못한 것들)</p>	<p>Q(1) 최대 우도 추정(Maximum Likelihood Estimation, MLE)에 대해서 이해가 잘 되지 않아 팀원 들과 함께 어떤것인지 조사해보고 논의해보았다.</p> <p>A(1) 우도는 우리가 알고 있는 표본 x들로 계산해 봤을 때, 추정했던 θ에 대한 추정이 얼마나 그럴듯 한지를 비교하는 척도 이다. $L(\theta)=f(x \theta)$로 계산할 수 있다. 우도가 클 수록 파라미터에 대한 추정이 그럴듯했다라는 의미가</p>

된다.

우리는 가지고 있는 데이터 x 로 추정한 파라미터가 그럴듯하기를 원하고, 그럼 최대 우도를 갖는 파라미터를 찾으면 된다. 그것이 최대 우도 추정이다.

간단하게 생각해 보면, 이항분포와 같은 그래프는 미분이 가능합니다. 그렇다면, 미분값이 0인 지점을 찾으면, 최대 우도를 갖는 파라미터를 바로 구할 수 있다. 하지만 미분이 불가능한 함수들도 존재한다. 이러한 경우는 다른 접근 방법들로 파라미터를 찾아나가야 한다.

마지막으로, 이러한 최대 우도 추정은 머신러닝 내부 파라미터를 최적화하는데도 자주 사용됩니다. 머신러닝은 대개 X 를 통해 y 를 출력하는 f 의 파라미터를 구하는 문제이기 때문에, 위에서 말한 우도 추정 문제와 굉장히 유사하죠. 흔히 얘기하는 Cross-entropy를 최소화하는 것과 우도를 최대화하는 것은 본질적으로 같다.

Q2

정보를 표현하는 방법과 계산하는 방법은 어떻게 있나요?

A2

어떠한 사건이 일어난 후, 또다른 사건이 일어나는 경우의 확률을 알아보기 위하여 베이지 정리, 최대 우도등을 활용합니다.

해당 사건들의 정보량을 표현하기 위해 엔트로피, 자기 정보라는 용어를 사용해 나타냅니다. 엔트로피는 어떠한 사건에 대하여 얼마나 예측이 불가능한지를 나타내고, 자기 정보는 해당 엔트로피에 의해 벌어진 사건은 얼마만큼의 정보량을 지닌지 표현할 수 있습니다.

어떠한 사건이 일어난 후, 또다른 사건이 일어나는 경우의 확률을 알아보기 위하여 베이지 정리, 최대 우도등을 활용합니다.

해당 사건들의 정보량을 표현하기 위해 엔트로피, 자기 정보라는 용어를 사용해 나타냅니다. 엔트로피는 어떠한 사건에 대하여 얼마나 예측이 불가능한지를 나타내고, 자기 정보는 해당 엔트로피에 의해 벌어진 사건은 얼마만큼의 정보량을 지닌지 표현할 수 있습니다.

자기 정보와 엔트로피는 서로 식으로 교환될 수 있고, 각각 이산확률분포, 연속 확률분포로 표현될 수 있습니다.

이때 2가지의 사건을 가지고 엔트로피를 비교할 경우를 교차 엔트로피라고 할 수 있습니다. 교차 엔트로피는 주요 엔트로피 P 와 비교할 엔트로피 Q 의 확률분포의 차이를 구해 계산합니다.

	<p>자기 정보와 엔트로피는 서로 식으로 교환될 수 있고, 각각 이산확률분포, 연속 확률분포로 표현될 수 있습니다.</p> <p>이때 2가지의 사건을 가지고 엔트로피를 비교할 경우를 교차 엔트로피라고 할 수 있습니다. 교차 엔트로피는 주요 엔트로피 P와 비교할 엔트로피 Q의 확률분포의 차이를 구해 계산합니다.</p> <p>Q3 키다이버전스는 무엇인가요?</p> <p>A3 쿨백 라이블러 발산은 두 확률분포의 차이를 계산하는데 사용하는 함수로 어떤 이상적인 분포에 대해 그 분포를 근사하는 다른 분포를 사용해 샘플링을 한다면 발생할 수 있는 정보 엔트로피 차이를 계산한다. 쉽게 말해 두 확률 분포의 차이를 계산하는데 사용하는 함수이다. 이산 확률 변수와 연속 확률변수에서 KL 다이버전스를 구하는 수식은 다음과 같다.</p> <p>"어떠한 확률분포 P가 있을 때, 샘플링 과정에서 확률분포 P를 사용할 경우의 엔트로피" 는 확률변수 P의 엔트로피, 즉 $H(p)$ 이다.</p> <p>또한, "어떠한 확률분포 P가 있을 때, 샘플링 과정에서 확률분포 Q를 P 대신 사용할 경우 엔트로피" 는 확률변수 P와 Q의 cross entropy, 즉 $H(p, q)$이다.</p> <p>키다이버전스의 특징으로는 두 가지가 있다. 첫번째는 '키다이버전스의 값은 0 이상이다'고 두번째는 '키다이버전스의 값은 asymmetric하다'이다. 우선, $H(p)$ 값이 최소 자원량의 기댓값이므로, $H(p)$는 무조건 $H(p, q)$보다 작을 수밖에 없다. 따라서 $H(p, q) - H(p)$ 인 키다이버전스의 값이 항상 0 이상인 것이다. 또한, p와 q의 위치가 바뀌면 키다이버전스의 값 또한 바뀌게 된다. 따라서 키다이버전스의 값은 symmetric 하지 않고, asymmetric 하다는 것을 확인할 수 있다.</p>
질문 내용	<p>1. 미분 불가능 함수에서는 Gradient descent 접근 방법으로 함수의 최대, 최소를 구하기가 힘듭니다.</p> <p>2. 미분 불가능 함수에서는 어떤 접근 방법들로 최대 우도를 구할 수 있는지 궁금합니다.</p>

기타	없습니다