Cross-Entropy Loss Function

크로스 엔트로피 손실함수

류영표 강사
youngpyoryu@dongguk.edu

강사
소개

류영표

Youngpyo Ryu

동국대학교 수학과/응용수학 석사수료

現 Upstage AI 부스트캠프 멘토

한국파스퇴르연구소 Image Mining 인턴(Deep learing)

前 ㈜셈웨어(수학컨텐츠, 데이터 분석 개발 및 연구인턴)

강의 경력

- 현대자동차 연구원 강의 (인공지능/머신러닝/딥러닝/강화학습)
- 딥러닝 집중 교육과정 강사
- (재)윌튼블록체인 6일 과정 (파이썬기초, 크롤링,머신러닝)
- 서울특별시 X AI 양재허브 X 모두의연구소 (중급 NLP과정) 보조강사
- SK아카데미_HLP(임원) 1차/2차 보조강사
- ㈜ 모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)
- LG전자 / LG 인화원 보조강사

주요 프로젝트 및 기타사항

- 제1회 인공지능(AI)기반 데이터사이언티스트 전문가 양성과정 최우수상 수상(Q&A 챗봇)
- 인공지능(AI)기반 데이터사이언티스트 전문가 양성과정 1기 수료
- 제 1회 산업 수학 스터디 그룹 (질병에 영향을 미치는 유전자 정보 분석)
- 제 4,5회 산업 수학 스터디 그룹 (피부암, 유방암 분류)
- 빅데이터 여름학교 참석 (혼잡도를 최소화하는 새로운 노선 건설 위치의 최적화 문제)

Loss function

퍼셉트론 및 선형회귀 분석에서 사용한 손실 함수

Square Loss : $$E(W) = \frac{1}{2} \sum_{d \in D} (y_d - \widehat{y_d})^2$$

y_d : Target

$\widehat{y_d}$: Output

Mean Square Loss : $$E(W) = \frac{1}{2} \frac{1}{|D|} \sum_{d \in D} (y_d - \widehat{y_d})^2$$

y_d : Target

$\widehat{y_d}$: Output

Information

<u>어떤 사건을 수치로</u> 나타낸 것.

확률을 이용한다.

확률사건 : Event

P(앞면)

P(앞면사건)

P(뒷면사건)

사건을 수치로 맵핑

$$\begin{cases} 앞면사건 \rightarrow 0 \\ 뒷면사건 \rightarrow 1 \end{cases}$$

$$P(X = 0)$$

$$P(X = 1)$$

확률 변수 : Random Variable

$$P(X = x) = P_x(x) = P(x)$$

다 같은 표현

Information

지금부터 사건을 수치화 해보자.

사건이 드물게 발생할 수록 정보가 커야 한다. $\frac{1}{P(x)}$

두 사건에 대한 정보는 곱해야 하나 더해야 하나? $\frac{1}{P(x)}$ $\frac{1}{P(y)}$

Information

$$\frac{1}{P(x)} \times \frac{1}{P(y)}$$

두 사건에 대한 수치를 더할 수 있도록 다시 생각해보자.

Information

지금부터 사건을 수치화 해보자. $$\frac{1}{P(x)}$$ 사건이 드물게 발생할 수록 정보가 커야한다.

두 사건에 대한 정보는
곱해야 하나 더해야 하나? $$\frac{1}{\mathrm{P(x)}} \times \frac{1}{\mathrm{P(y)}}$$ 정보로 생각하면 더하는 것이 직관적이다.

$$log\frac{1}{\mathrm{P(x)}} + log\frac{1}{\mathrm{P(y)}}$$

로그는 고맙게도 곱셈을 덧셈으로 변경해준다.

Information

$$log\,\frac{1}{\mathrm{P(x)}}$$

$$logP(x)^{-1}$$

$$-logP(x)$$

Entropy

정보량의 기댓값

확률 변수 X의 기댓값 계산 공식 $E[X] = \sum_x xP(x)$ $E[X] = \int x\,P(x)\,dx$

확률변수 X의 분포함수

확률질량함수 (Probability mass function)

확률밀도함수 (Probability density function)

Entropy

정보량의 기댓값

확률 변수 X의 기댓값 계산 공식

$$E[X] = \sum_{x} xP(x)$$

확률변수 X의 분포함수가 중요

$$E(aX + b) = \sum_{x} (ax + b)\, P(x)$$

$$E[f(X)] = \sum_{x} f(x)\, P(x)$$

일반화 시켜보면 ?

Entropy

정보량의 기댓값

$$E[X] = \sum_{x} xP(x)$$

확률변수 X의 분포함수가 중요

$$E[-logP(x) \quad] = \sum_{x} -logP(x) \quad P(x)$$

Entropy

$$\sum_{x} -logP(x) \quad P(x)$$

일반화 시켜보면 ?

Cross-Entropy

다른 확률을 곱해서 Entropy를 계산한 것.

예를 들어, 0또는 1만 가지는 확률변수 X에 대해

$$\mathrm{Entropy} = -\mathrm{p}(X = 0)\log\big(\mathrm{p}(x = 0)\big) - \mathrm{p}(X = 1)\log(\mathrm{p}(X = 1))$$

교차 : Cross

$$\text{Cross-Entropy} = -\mathrm{p}(X = 0)\log\big(\mathrm{q}(x = 0)\big) - \mathrm{p}(X = 1)\log(\mathrm{q}(X = 1))$$

★ 참고로 0,1 만 가지는 특별한 확률변수는 이름이 있다. 베르누이: Bernoulli

신경망의 손실함수로 활용

다른 확률을 곱해서 Entropy를 계산한 것.

$$E(W) = \frac{1}{2}\frac{1}{|D|}\sum_{d \in D}(y_d - \widehat{y_d})^2$$

Cross-Entropy Loss

$$E(W) = \frac{1}{2}\frac{1}{|D|}\sum_{d \in D}(y_d \log(\widehat{y_d}) - (1 - y_d)\log(1 - \widehat{y_d})$$

0~1 사이

0 또는 1

표본으로부터 Bernoulli 분포에 대한 파라미터를 추정
Maximum likelihood estimation

참고

- 딥러닝을 위한 신경망 기초 (고재필 교수님 강의안 참고)

Thank you.

크로스 엔트로피 손실함수 실습

류영표 강사 / youngpyoryu@dongguk.edu