

17. 몬테카를로 방법

무작위 알고리즘

- 1) 라스베이거스 알고리즘 : 정확한 답 or 답을 찾지 못함
- 2) 몬테카를로 알고리즘 : 무작위한 오차가 있는 답을 리턴

기계학습 문제 -> 정확한 답을 기대할 수 없음

→ 결정론적 근사 알고리즘 혹은 몬테카를로 근사 방법 사용

17.1 표본추출과 몬테카를로 방법

표본추출을 사용하는 이유

1. 낮은 비용
2. 처리 불가능한 수준의 합이나 적분 근사 가능

몬테카를로 표집 :

합&적분을 어떤 분포의 기댓값으로 간주, 평균을 통해 그 기댓값을 근사

$$s = \sum_{\mathbf{x}} p(\mathbf{x}) f(\mathbf{x}) = E_p[f(\mathbf{x})]$$

$$s = \int p(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = E_p[f(\mathbf{x})]$$

17.1 표본추출과 몬테카를로 방법

몬테카를로 추정량 : s 를 n 개의 표본에 대한 평균으로 근사

이러한 근사가 정당한 이유

1. unbiased estimator
2. 큰 수의 법칙

$$\hat{s}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}^{(i)}).$$

17.2 중요도 표집

몬테카를로 방법에 쓰이는 적분 또는 합산을 분해하는 대신

$$p(\mathbf{x})f(\mathbf{x}) = q(\mathbf{x}) \frac{p(\mathbf{x})f(\mathbf{x})}{q(\mathbf{x})}.$$

q 에서 표본을 추출해 $p(\mathbf{x})f(\mathbf{x})/q(\mathbf{x})$ 의 평균을 구하는 방법

→ 최적 표집 분포 함수 q 는 **최적 중요도 표집**을 사용해 구함

17.2 중요도 표집

몬테카를로 추정량 -> 표집 추정량으로 변환 가능

$$\hat{s}_q = \frac{1}{n} \sum_{i=1, \mathbf{x}^{(i)} \sim q}^n \frac{p(\mathbf{x}^{(i)})f(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})}$$

(최적의 분포를 사용한다면 하나의 표본만을 사용할 수 있음)

최적의 분포(기댓값의 분산이 최소) : q^* 라고 표기

17.2 중요도 표집

또 다른 방법 : 편향 중요도 표집

- p나 q를 정규화할 필요 x

$$\frac{\sum_{i=1}^n \frac{\tilde{p}(\mathbf{x}^{(i)})}{\tilde{q}(\mathbf{x}^{(i)})} f(\mathbf{x}^{(i)})}{\sum_{i=1}^n \frac{\tilde{p}(\mathbf{x}^{(i)})}{\tilde{q}(\mathbf{x}^{(i)})}}.$$

q를 잘못 선택 -> 효율성이 크게 나빠짐

q를 잘 선택하여 몬테카를로 추정을 사용하는 예

- 1) 어휘가 큰 신경망 언어 모형의 훈련
- 2) 분배함수 추정
- 3) 자동부호기와 같은 심층 모형

17.3 마르코프 연쇄 몬테카를로 방법

기본적인 보장 : 모형이 그 어떤 상태에도 0의 확률을 배정하지 않음

에너지 기반 모형(EBM)에서 표본추출이 어려운 이유 :

$p(a, b)$ 중 a 의 표본을 추출하기 위해 $p(a|b)$ 에서 a 를 뽑아야 함

b 의 표본을 추출하려면 $p(b|a)$ 에서 표본을 추출해야 함

마르코프 연쇄의 핵심 : 상태 x 의 값을 임의의 값으로 초기화 후 무작위로 갱신하는 과정 반복

17.3 마르코프 연쇄 몬테카를로 방법

어떤 확률변수가 가질 수 있는 상태의 수가 가산적이라고 가정
-> 양의 정수로 표현 가능

연쇄 시작 지점 : 어떤 분포 $q(0)$ 을 이용하여 각 x 를 임의의 값으로 초기화

목표 : $q(x)$ 가 $p(x)$ 에 수렴하게 하는 것

17.3 마르코프 연쇄 몬테카를로 방법

$$(1) \quad q(x = i) = v_i.$$

$$(2) \quad q^{(t+1)}(x') = \sum_x q^{(t)}(x) T(x' | x). \text{ 주어진 } x \text{가 } x' \text{로 전이할 확률}$$

$$(3) \quad A_{i,j} = T(\mathbf{x}' = i | \mathbf{x} = j).$$

$$(4) \quad \mathbf{v}^{(t)} = \mathbf{A} \mathbf{v}^{(t-1)}.$$

$$(5) \quad \mathbf{v}^{(t)} = \mathbf{A}^t \mathbf{v}^{(0)}.$$

17.3 마르코프 연쇄 몬테카를로 방법

연소 : 평형분포에 도달할 때까지 마르코프 연쇄를 실행하는 것

-> 무한히 많은 표본 추출 가능

- 유한한 개수의 표본들은 평형분포를 잘 대표하지 못할 수 있음

-> 매 n 개의 표본마다 하나의 표본을 돌려주어 해결

-> 이웃한 표본 사이의 큰 상관관계 방지

-> 전체적인 계산 비용 증가

17.4 기브스 표집

분포 $q(x)$ 를 잘 뽑았는지 확인할 때

- 1) 학습된 분포 p 로부터 T 유도
- 2) T 를 직접 매개변수화하여 해당 분포 p 을 정의할 수 있도록
그 매개변수들을 학습

기브스 표집을 이용하면 마르코프 연쇄를 간단하고 효과적으로 구축 가능

기브스 표집 : 하나의 변수 x 선택, x 의 이웃 변수들을 조건으로 하여 $p(x)$ 로부터 그 변수의 표본 추출

17.5 분리된 모드 사이의 혼합과 관련된 어려움들

마르코프 연쇄 몬테카를로 방법의 어려움

- 마르코프 연쇄가 잘 혼합되지 않음
- 표본들 사이 상관관계가 아주 높을 수 있음
- > '혼합이 느리다'

마르코프 연쇄는 더 낮은 에너지 구성이 산출되는 방향 선호
에너지가 낮은 영역 -> mode

17.5 분리된 모드 사이의 혼합과 관련된 어려움들

고확률 피크점, 주위 저확률 영역들이 있는 분포

-> 고확률 영역들을 낮추고 저확률 영역을 높이기

이러한 기법에 기초한 마르코프 연쇄

-> 고확률 분포에서 표본들을 뽑아 서로 다른 모드들을 혼합한 후
단위 분포의 표집을 재개

감사합니다