1 베이지안 네트워크

1 베이지안 정리

- 두 사건 A와 B에 대해서 B가 발생한 후에 사건 A가 발생할 확률을 계산하면 다음과 같은 식을 유도
- 사전 조건이 없는 상태에서 사건 A가 발생할 확률 P(A)를 알고 어떤 조건 B를 전제로 한 A의 발생 확률 P(A|B)를 구하는 경우

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

1 베이지안 정리

- 베이즈 확률 적용 시 분모가 직접 주어지지 않은 경우,
  - 조건부 확률을 이용하여 계산

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

  - 일반화 하여,

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

1 베이지안 네트워크

1 베이지안 정리

베이지안 정리의 특징

- 명확한 사전 확률 요구
- 각 사건들은 상호 배타적이며 독립적

베이지안 정리의 단점

- 실세계에서는 사전 확률을 명확히 유지하기 힘듦
- 각 사건들의 연관성으로 인하여 상호 배타적이며 독립적이기 어려움
- 잘 정의된 좁은 영역의 문제 해결에 유용함

1 베이지안 정리

베이즈 정리의 의미

- 증거에 따른 가설의 정확도
  - A를 가설로 간주하고 B를 증거로 간주하는 상황
  - 처음에는 가설 A에 대한 사전 확률 정보만 있지만 증거 B가 주어짐에 따라 정확도가 개선된 가설 A에 대한 사후 확률 $P(A|B)$를 베이즈 정리로 계산 가능
  - 연속된 사건이 발생할 경우 다음 확률의 계산이 가능하다는 의미

$$P(\text{가설}|\text{증거})$$

1 베이지안 정리

베이즈 정리의 의미

- 증상을 보고 원인을 추정하는 진단을 하는 경우
  - A를 원인으로 간주하고 B를 증상들로 간주하는 상황
  - 사전 확률 $P(A)$는 고장이나 질환이 생겼을 때 다른 정보가 없는 상태에서 A가 원인일 확률
  - 가능도 $P(B|A)$는 A가 원인일 때 B의 증상이 나타날 확률
  - 사후 확률 $P(A|B)$는 B의 증상이 나타날 때 원인이 A라고 판정할 확신도

  $P(\text{원인}|\text{증상})$

1 베이지안 정리

베이즈 정리의 의미

- 기계 학습에 적용하는 경우
  - 학습 데이터를 D, 모델은 파라미터를 의미하는 θ로 정의

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

P(모델의 파라미터|학습데이터)

  - 학습데이터를 나타내는 패턴에 대응하는 모델의 형태를 결정하는 파라미터를 찾기 위한 기본적 이론

1 베이지안 정리

베이즈 정리의 의미

- 기계 학습에 적용하는 경우
  - 점진적 기계 학습을 가능하게 하는 프레임을 제공
  - 모델의 파라미터에 대한 사전 확률과 학습 데이터를 사용하여 모델의 파라미터에 대한 개선된 사후 확률을 결정하는 역할
  - 모델 파라미터에 대한 사후 확률을 가지고 있는 상태에서 추가적인 학습 데이터가 주어지면 이미 가지고 있는 사후 확률을 사전 확률로 가정하여 베이즈 정리를 새로 추가된 학습데이터에 적용하여 보다 개선된 확률 정보를 획득

2 베이지안 네트워크

- 전문가 시스템은 주어진 조건에 적합한 답변을 하지만 추론 규칙이 정교하지 않음

정의

- 확률 개념을 도입해 추론 규칙을 개선한 전문가 시스템으로 제안한 것

2 베이지안 네트워크

특징

- 불확실성을 포함한 사건의 예측과 관측 결과를 활용
- 장애 진단에 사용하는 그래픽 확률 모델
- 각 노드는 확률 변수
- 확률 변수 사이의 확률 의존 관계 정보를 유향 그래프로 나타내는 네트워크로 시스템으로 구성

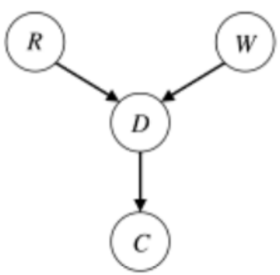
2 베이지안 네트워크

베이지안 네트워크(Bayesian Network)

- ‘빌리프 네트워크(Belief network)’라고도 불림
- 집합을 조건부 독립으로 표현하는 확률의 그래픽 모델
- 추론과 학습을 수행하기 위한 효과적인 알고리즘이 존재
- 예를 들어, 질환과 증상 사이의 확률 관계를 나타낼 수 있음
- 증상이 주어지면 다양한 질병의 존재 확률 계산 가능

2 베이지안 네트워크

- 유향 그래프 내에 인접한 노드 사이에는 조건부 확률 테이블이 할당
  - 은닉 마르코프 모델과 유사



R: 비가 온다

R	*P*(*R*)
0	0.8
1	0.2

W: 바람이 강하다

W	*P*(*W*)
0	0.7
1	0.4

D: 전철이 지연된다

		P(*D*|*R*, *E*)	
R	*E*	*D* = 0	*D* = 1
0	0	0.8	0.2
0	1	0.9	0.1
1	0	0.2	0.8
1	1	0.3	0.7

C: 지각한다

	P(*C*|*D*)	
D	*C* = 0	*C* = 1
0	0.9	0.1
1	0	1

[베이즈 네트워크와 조건부 확률 테이블]

R: 비가 온다

W: 바람이 강하다

D: 전철이 지연된다

C: 지각 한다

[비가 내리고 바람이 부는 날에 지각할 확률 계산]

2 베이지안 네트워크

단점

- 네트워크가 복잡해 질 수록 조건부 확률 테이블 역시 복잡해 짐
- 일반적인 네트워크 구조로는 확률 추론이 어려워 다양한 방법을 사용하여 사후 확률을 구해야 함
- 무향 그래프이면서 루프가 없는 단일 결합 네트워크의 경우
  - 베이즈 정리를 이용하여 비교적 쉽게 임의의 사후 확률 산출 가능
- 여러 개의 결합 네트워크의 경우
  - 확률 계산이 복잡해져 계산 비용 증가

2 베이지안 네트워크

베이지안 네트워크의 단점 보완

- 사후 확률을 구하는 효율적인 방법 연구
  - 사전에 단일 결합 트리 구조 그래프로 변환
  - 정밀도를 높여 다양한 샘플링 기법을 이용해 차이를 줄이는 방법
- 노이즈 데이터가 포함되어 불확실성이 있는 상황
  - 센서 등으로 관측한 결과를 기반으로 진단과 인식 등의 추론 도구를 이용

2 베이지안 분류기

1 베이지안 분류기

나이브(단순) 베이즈 분류기

- 기계학습에서 데이터의 속성들 사이의 독립 사건을 가정하는 베이즈 정리를 적용한 확률 분류기의 일종
- 특징
  - 지도 학습에 효율적
  - 모델 파라미터 추정 시 최대 가능도 법 을 사용
  - 파라미터 추정에 사용되는 학습 데이터 양이 많이 필요하지 않음
  - 간단한 디자인과 단순한 가정이 실제 응용에 효율적으로 동작

1 베이지안 분류기

학습 규칙

- 최대가능도법
  - 임의의 데이터가 얻어질 확률 $P(x)$를 가정
  - $P(x)$를 최대화하는 미지의 파라미터 w를 결정
  - w 는 하나(벡터 또는 변수)가 결정

1 베이지안 분류기

베이즈 분류기

- w가 확률적으로 결정
- 확률을 업데이트 하는 학습을 수행
- 확률을 업데이트 하기 위해 베이즈 정리를 적용

2 베이지안 분류기 개념

- $p(C_k|x_1,\ldots,x_n)$ 입력 특징을 벡터화 하면,
- $p(C_k|\mathbf{x}) = \dfrac{p(C_k)P(\mathbf{x}|C_k)}{p(\mathbf{x})}$ 분자 부분은 결합 확률이므로,
- $p(C_k,x_1,\ldots,x_n)$ 로 표현되며, 연쇄 법칙을 사용하면,

$$\begin{aligned} p(C_k,x_1,\ldots,x_n) &= p(C_k)\ p(x_1,\ldots,x_n|C_k) \\ &= p(C_k)\ p(x_1|C_k)\ p(x_2,\ldots,x_n|C_k,x_1) \\ &= p(C_k)\ p(x_1|C_k)\ p(x_2|C_k,x_1)\ p(x_3,\ldots,x_n|C_k,x_1,x_2) \\ &= p(C_k)\ p(x_1|C_k)\ p(x_2|C_k,x_1)\ \ldots p(x_n|C_k,x_1,x_2,x_3,\ldots,x_{n-1}) \end{aligned}$$

- 각 사건이 독립적이라고 가정하면

$$p(C_k|x_1,\ldots,x_n) = \frac{1}{Z}p(C_k)\prod_{i=1}^{n} p(x_i|C_k)$$

베이즈정리

2 베이지안 분류기 개념

$$\hat{y} = argmaxP(C_k)\prod_{i=1}^{n} P(x_i|C_k)$$

$$k \in \{1, \dots, K\}$$

2 베이지안 분류기 개념

파라미터 추정

- 클래스 사전 확률
  - 클래스 간의 동일한 확률을 가정하여 산출
  - 학습 데이터로부터 클래스 확률 추정치 산출
- 특징의 분포
  - 분포를 가정
  - 가우시안 분포, 베르누이 분포, 다항 분포 등

3 베이지안 분류기 학습

파라미터 추정

평균과 분산은 특정 데이터에 따라 학습이 되며, 데이터가 많을 수록,
특정 데이터가 전체 데이터의 분포를 따를 수록 학습 정확도가 높아진다.

- 데이터 집합
  - 평균

$$\frac{dL}{d\mu} = \sum_t \frac{(x^t - \mu)^2}{\sigma^2} = 0 \qquad \sum_t x^t = \sum_t \mu \qquad \mu(or\ m) = \frac{\sum_t x^t}{N}$$

  - 분산

$$\frac{dL}{d\sigma^2} = \sum_t [-\frac{1}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{(x^t - \mu)^2}{(\sigma^2)^2}] = 0 \qquad \sum_t \sigma^2 = \sum_t (x^t - \mu)^2 \qquad \sigma^2 = \frac{\sum_t (x^t - \mu)^2}{N} = \frac{\sum_t (x^t)^2}{N} - m^2$$

[데이터 집합으로부터 산출된 각 변수의 평균과 분산 식]

3 베이지안 분류기 학습

- 앞서 학습한 결과 각 입력 변수에 따른 평균과 분산이 산출
- 2진 분류를 하고자 할 경우, 두 클래스에 속할 확률 중 큰 클래스를 선택

$$\boxed{arg\max_{k} P(C_k|x)}$$

①
$$P(x|C_1)P(C_1) = P(x_1|C_1)P(x_2|C_1)P(C_1)$$
$$\frac{1}{\sqrt{2\pi s_{11}}}e(-\frac{1}{2}\frac{(x_1-m_{11})^2}{s_{11}{}^2}) * \frac{1}{\sqrt{2\pi s_{12}}}e\left(-\frac{1}{2}\frac{(x_2-m_{12})^2}{s_{12}{}^2}\right) * P(C_1)$$

②
$$P(x|C_2)P(C_2) = P(x_1|C_2)P(x_2|C_2)P(C_2)$$
$$\frac{1}{\sqrt{2\pi s_{21}}}e(-\frac{1}{2}\frac{(x_1-m_{21})^2}{s_{21}{}^2}) * \frac{1}{\sqrt{2\pi s_{22}}}e\left(-\frac{1}{2}\frac{(x_2-m_{22})^2}{s_{22}{}^2}\right) * P(C_2)$$

[각 변수의 사전 확률을 이용하여 산출된 클래스 별 포함 확률]

3 확률 모델을 이용한 베이지안 분류기 생성

1 베이지안 분류기 예 1

▶ 키, 몸무게, 신발사이즈를 이용한 성별 판단

[데이터]

성별	키	몸무게	신발사이즈
남	182.9	82	30
남	180.4	86	28
남	170.1	77	30
남	180.4	75	25
여	152.4	45	15
여	167.6	68	20
여	165.2	59	18
여	175.3	68	23

[쿼리 데이터]

성별	키	몸무게	신발사이즈
?	182.9	60	20

△ 쿼리 데이터를 이용하여 성별을 판단

◁ 8명의 데이터의 각 특징에 대한 평균과 분산을 산출

1 베이지안 분류기 예 1

분모는 동일하며 단지 스케일링 역할만 하므로 무시

▶ 클래스 사전 확률 및 변수 별 사전 확률 산출

$$P(male)=(P(male)P(height|male)p(weight|male)p(footsize|male))/evidence$$
$$P(female)=(P(female)P(height|female)p(weight|female)p(footsize|female))/evidence$$
$$evidence=P(male)P(height|male)p(weight|male)p(footsize|male)+P(female)P(height|female)p(weight|female)p(footsize|female)$$

- 클래스 사전 확률인 $P(male)$과 $P(female)$은 0.5로 가정하거나, 특정 통계 수치를 이용한 사전 확률로 사용
- 입력 데이터의 확률 분포는 가우시안 분포를 가정

1 베이지안 분류기 예 1

- 사후 확률 산출
  - 사전에 산출된 데이터 집합에서의 각 변수 별 평균과 분산 이용
  - 클래스 별 사후 확률을 산출

P(height|male)	남자일 때 키의 확률 데이터		0.736523326	14.26284	0.051639
P(weight|male)	남자일 때 몸무게의 확률 데이터		0.000301088	12.44931	2.42E-05
P(footsize|male)	남자일 때 신발크기의 확률 데이터		0.002253774	5.922932	0.000381
사후 확률(male)		=	**2.37614E-10**	◁ 쿼리 데이터가 남자일 확률	
P(height|female)	여자일 때 키의 확률 데이터		0.174641291	23.84977	0.007323
P(weight|female)	여자일 때 몸무게의 확률 데이터		1	27.22895	0.036726
P(footsize|female)	여자일 때 신발크기의 확률 데이터		0.956841381	8.438568	0.113389
사후 확률(female)		=	**1.52466E-05**	◁ 쿼리 데이터가 여자일 확률	

2 베이지안 분류기 예 2

▶ 문서 분류 예

문서번호	주요단어	문서분류
1	fun, couple, love, love	comedy
2	fast, furious, shoot	action
3	couple, fly, fast, fun, fun	comedy
4	furious, shoot, shoot, fun	action
5	fly, fast, shoot, love	action

* 입력 데이터
fast, furious, fun

◁ 5개의 문서에 포함된 단어를 확률로 산출

◁ 다음과 같은 입력 데이터가 있는 문서를 어떤 문서로 분류할 것인가

2 베이지안 분류기 예 2

- 사후 문서 분류 확률 산출
  - Comedy 문서일 확률

$$P(comedy|x)=p(comedy)p(fast|comedy)p(furious|comedy)p(fun|comedy)$$

$$=\frac{2}{5}\frac{1}{9}\frac{0}{9}\frac{3}{9}=0$$

  - Action 문서일 확률

$$p(action|x)=p(action)p(fast|actino)p(furious|action)p(fun|action)$$

$$=\frac{3}{5}\frac{2}{11}\frac{2}{11}\frac{1}{11}=0.0018$$

3 베이지안 분류기의 문제점

- 사전 조건부 확률이 0인 경우에는 어떤 학습 데이터가 들어오더라도 사후 확률이 0이 아닌 값을 가질 수 없음

- 해결방법
  - 모든 속성값과 클래스 조합에 대한 빈도에 임의의 수를 더해 계산을 수행하도록 함

3 베이지안 분류기의 문제점

- 입력 데이터가 실수일 경우 입력 데이터에서 각 입력 변수의 개수를 세는 부분이 불가, 사전 확률 산출의 문제 발생

$$P(x = v|C_k) \Rightarrow {}^{1}/_{C_k}$$

- 해결방법
  - 입력 변수의 구간을 임의로 구간별로 묶는 방법 사용