### < [Lecture12 (03/30) : 베이지안 결정 이론(Bayesian Decision Theory)](https://nbviewer.jupyter.org/github/jeonghunyoon/machine-learning-lecture-notes/blob/master/Lecture12_Bayesian_Decision_Thoery.pdf)>

머신 러닝의 정의 정도는 외울 것(탐 미첼)-> 암기암기암기

통계 모델링과 머신 러닝의 차이점은 선형회귀 모형때 확실히 느낄 수 있을 것임.

손실 함수(알고리즘을 다뤄 보면서 언제 손실 함수를 쓰는지 알아보는 게 좋음)-> 하지만 손실함수의 정의는 알아둘 것

비지도 학습(언 슈퍼바이즈드 학습)

퍼뮤테이션을 이용해서 아노바를 하는 방법

-나이브 베이지안

:1) 직접 코드 써서 구현 하는 방법 2) 라이브러리를 이용해서 구현 하는 방법

텀스에 용어는 다 알고 갈 것!!!

디시젼 띠오리!!!!!!!

라플라시안 스무딩(데이터가 없는 경우, 어떻게 고정해주는 지)

-랜덤 베리에이블

1. 엑스라는 함수에 씨라는 사건을 넣으면 맵핑되는 수! 여기서 함수 엑스가 랜덤 베리에이블임

랜덤 베리에이블의 정의는 함수일뿐임. 사건을 실수에 대응시켜주는 함수가 랜덤 배리에이블임!(엑스)

-베이시아 룰

: 파란 공이 관찰된 것임. 파란 공은 관찰 값이고 주머니 에이 또는 비에서 나오는 것은 관찰 이전의 사건임.

우리의 사고를 수학적으로 표현하면 조건부 확률이 됨!.

베이시아 룰 - 대답 1) 우도 함수/우도 (우도 확률 아님)

대답 2)

사전 확률 ( 사전 함수 아님. 우도 함수랑 비교할 것)

파랑 공을 뽑기 전의 사건이기 때문에 사전 확률 임!

파랑 공이 주어졌을때, 먼저 일어난 사건 어디에서 일어났을 까-> 그래서-> 사후 확률이라고 함.

-베이즈 룰의 수학적 유도

조건으로 주어지는 게 뒤에 일어나는 사건(뒤에 일어나는 사건: 파란공)

결합확률( 엑스 교집합 와이나 와이 교집합 엑스는 같을 것)

와이가 나중에 발생할 사건이고 엑스는 먼저 일어난 사건이다!

전(토탈)사건의 확률

분류를 할 때 에이에 속하든 비에 속하든 두개의 확률을 비교할 때 분모가 같을 때 분자만 가지고 비교하면 됨!

베이즈 룰 수학정 정의로 인해 3가지 인사이트가 나올 수 있음

우리가 관심이 있는 것은 사후 확률임!

베이즈 룰에서 우리가 관심이 있는 것은 어떤 관찰 값이 있을 때 그 앞에 발생했던 사건을 알고 싶은 것!-> 앞에 있는 사건을 추적하고 싶은 것!

세터는 우리가 구하고자 하는 모수가 될 것 그리고 엑스는 데이터(머신러닝 기초에서 배운 인풋 임)가 될 것/ 스패머의 모수를 결정 짓는 / 데이터가 주어졌을 때, 그 데이터를 생성시킨 파라미터(모수)를 찾아 나갈 것.(이게 오늘의 목표임)

우도 값은 관찰 값 자체이므로 찾아내기 쉬움.( 우도는 앞에 사건을 이미 알고 그 사건에 의해 발생한 사건 값임)

사후 확률은 사전확률과 우도 값을 혼합해서 알아내는 것임.(스팸메일분류가 여기에 해당)

분류기는 함수임!// 클래스는 부류 (분류 아님!// 분류기는 클래시파이어임)

-디시젼 띠오리

디시전 이론도 부류랑 비슷 함.

이미지를 벡터로 바꾸는 것은 메져먼트라고 얘기함!-> 디시젼 띠오리에서 나온 얘기임

그 매져먼트가 디시전 룰에 의해서 나눠짐

수치화! 메져먼트가 있을 때, 룰에 의해 분류하는 데

엑스가 들어왔을 때, 룰에 의하여 구분 되는 데, 디시젼 룰을 확률로 이용하는 것.

엑스가 씨1원에서 일어날 확률

엑스가 씨2에서 일어날 확률

* 어떤 엑스레이 사진이 암인지 암이 아닌지 구분 할 때, 특정 카테고리에 있는 사람이 있을 것이고 그 사람이 엑스레이를 찍을 것
* 내가 원하는 것은 사진을 보고 암에 걸렸는 지 안걸렸는지 확인 하는 것
* 그 사진은 어떤 사람이 찍은 것
* 사진이 존재 하는 것은 암에 걸린 사람 그룹 혹은 안걸린 사람 그룹내에서 사람을 봅아서 사진을 찍는 것
* 그래서 결합 확률 분포 을 쓰는 것/ 사전 확률!!!!!!!!!
* 선후 관계를 이해 할 것 ( 그리고 디시젼 띠오리의 수학 식을 이해하는 것이 중요함)
* 디시젼 띠오리를 이용해서 부류

전체 공간은 입력 벡터가 존재하는 공간이고 결정 경계를 기준으로 나눠진 하나하나의 공간이 결정 구역임.

결정경계는 우리가 보고 결정하는 것임 (결합 확률 밀도 함수를 가지고 결정경계를 구하는 것임)

사전 확률을 구하기는 어려움! ( 스패머 들의 모든 메일을 구하는 것은 사실 어려움!)

분포라는 게 변수 들이 찍히는 함수를 의미함!!!

사전확률은 임의로 설정함.

일주일 동안 들어온 메일 중에서 스팸메일의 비율은 알 수 있지만

전세계 스패머들의 전체 메일을 구할 수는 없음!!!!!!!-> 그래서 사전 확률은 구할 수 없음

추정(인터런스)을 잘하면 잘할수록 값은 정확해 질 것!

베이시안은 데이터를 계속 넣고 업데이트 하는 것.(데이터를 넣으면 넣을수록 정교해짐!)

* 결합 확률 밀도 함수!!!!!

베이시아 클래스(분류기)의 2가지 기준은 반드시 알아 둘 것

* 디시젼 바운더리 포 에버리지 에러

클래스는 두새만 있음. 엑스가 클래스 1 혹은 클래스2 에만 속하는 확률

엑스가 사진이고 사진은 몇차원 이겠지만 1차원으로 가정하는 것임.

그래프를 해석할 수 있어야 함.

분포 그래프 임!

결정경계를 어떻게 나눌 까가 가장 중요함.

결정결계는 에러를 최소화 하는 것임.

결정 경계 그래프에 의하면 잘 못 분류할 수 있음!( 실제 값과 분류 값의 차이)

그래프 상 색깔 표시는 잘못 분류할 확률 임.

결정 경계를 기준으로 해서 씨1으로 분류 해야하는 데 씨2로 분류 되거나 반대의 경우도 있음.

결정경계를 엑스 0로 하는 경우,

결정 경계에 따라서 오분류 되는 범위가 왔다갔다함.

결정결계 오류가 최소화 되는 것을 찾아야함.!

* 베이시안 디시젼 룰(에러)을 이용해서 최소화 되는 조건 찾기!
* 엑스 교집합 씨 1은 구하기 어려움 그래서 구하기 쉽게 하기 위해 조건부 확률로 구해야함.
* 추정을 할 수 있다 (엑스가 있는데 클래스가 있을 확률)
* 그래서 사후 확률로 다시 돌아가는 것임. ( 베이시아 디시젼 에러의 좋은 책은 강의 노트 순으로 설명함)
* 다분류로 가능 경우, 결합 확률 밀도 함수에서 겹쳐 지는 포인트가 여러 개일 수 있음. 그럴 때는 결정 경계를 여러 개 둠!/ 구간 마다 결정 경계를 여러 개 둠
* 리젝트 옵션

뜨레드 홀드를 줌(확실하지 않은 영역에 대해서는 분류하지 않는다.)

자동 제재하기 때문에 잘못 제재할 수도 있음.

0.5?????????

분류를 하지 않으면 정상으로 일단 두고 다른 제재 수단으로 다시 분류함.

사후 확률을 직접 구할 수 없으니 라이클리 후드랑 프라이어를 통해서 구함

-베이즈 클래시파이어

베이즈 공식 써주면 빨간색 결과값이 나옴.

최종 알고리즘

사용하기 좋은 것은 빨간색 공식임/ 라이클리 후드 사용함/ 사전 분포는 쓰기 어려움.

뉴머릭한 데이터를 다룰 때, 분포식에 숫자를 넣어서 구하고 싶을때는 가우시안 클래스를 사용함.(가우시안 클래스는 )/지금 단계에서는

라이클리 후드는 발생빈도를 보고 도수 분포표로 구해도 아무런 문제가 없다.!

-손실 함수

확률로 구분하면 되나, 실제 병에 걸렸으나 안걸렸다고 하는 게 정말 나쁜 거임.

잘못 분류 된 것 중에서 어떻게 잘못 분류되었는지를 보는 것

분류 된 것중에서도 잘못 분류 된 것 잘 못분류 된 것을 분류화 하는 것

기대 손실 최소화 버전

암인데 정상으로 보는 경우 1000점 주는 것.

-기대 손실의 최소화 버전

계수가 하나 더 붙음( 실제로는 제이 클래스에 속했는데 선택된 액션은 아이 일 때)

엑스를 암으로 분류 할 확률과 암에

씨제이는 정상일 확률 임.

엑스가 정상일 확률에서 계수를 더해줌 ( 실제로 제이인데 아이로 분류했을 때를 계산하여 계수로서 더해줌)

베시스 클래시파이어(기대손실의 최소화버전)을 더 많이 씀!

리스크를 최소화 하는 액션으로 선택하는 게 베이즈 클래시 파이어의 알고리즘임!

-타입원 에러인 이유는? 정상을 암으로 구분하는 것이 (타입원 에러의 정의는 귀무가설 참인데 귀무가설을 에러로 두는 것./정상인을 정상인으로 하는 것이 귀무 가설임. 정상인을 암으로 하는 경우에 사회적 비용(치료 비용)이 더 큼 그래서 타입 원이라고 함.

:로스 매트릭스는 위치는 정확하게 이해하는 것이 좋을 듯 하다.

:웨이트를 정하는 방법은 실제 데이터에서 웨이트를 정하는 방법은 1) 데이터 비율을 살살 볼 것 (그리고 데이터 비율에 잘못 분류하는 경우의 비용을 곱해서 웨이트를 결정할 것)

:기대 손실의 최소값에서 L11 은 손실 매트릭스라고 부름.

오분류 비율과 어떻게 두개의 클래스로 구분하는 지

-나이브 베이지안 클래시파이어 (\*\*\*\*\*\*)

보통 주변 우도는 구하지 않고 우도와 사전확률만 구함.

각각의 용어 중요함!

우도를 구하는 것은 쉽지 않음-> 그이유는?->1) 계산 수가 굉장히 방대해서 구하기 어렵다.-> 그래서 우도를 구하는 쉬운 버전은 컨디셔널 독립임!

컨디셔널 독립

조건부 독립의 가정이기 때문에, 번개(제트)가 일어났을 때 천둥과 비가 나왔을 확률은 번번개 일어났을 때 천둥이 나올 확률 곱하기 번개가 일어났을 때 비가 나올 확률과 같다.

( 파워 풀한 분류기 이지만, 어순을 바꿨을때도 적용 되므로 주의를 기울 여야 한다.)

-파라미터 인 나이브 베이즈

여기서 티와 엘이 1인 확률을 빼는 이유는

조건부 독립을 사용하는 이유는 파라미터의 수를 줄이기 위해서임.

그럼에도 불구하고 성능은 좋다.

-나이브 베이스 클래시파이어의 분류 조건

에서 조건부 독립을 따로 구분하는 것이 비용이 더 적게 든다!

엘에스 이\_각 문장의 길이는 코퍼스의 길이임.\_ 이때의 코퍼스와 나이브 클래스파이어의 코퍼스는 같은 것임.

아규먼트 맥스 와이 란? 1) 와이가 스팸일 때 조건부 독립 계산 =에이

2) 와이가 햄일 때 조건부 독립 계산= 비

일 때 에이와 비 중 큰 값을 만드는 와이의 값을 반환 하라.

* 나이브 베이스 클래스파이어 예제

여기서 코퍼스는 3개 임.( 복권/백만/수신취소)

백 오브 워즈 벡터는 전체 단어 셋(코퍼스)를 이야기 함.\_자연어 처리

라이클리 후드(우도) 란 스팸일 때 워드가 나올 확률\_피(스팸)은 피(스팸|워드)

-라이클리 후드랑 프로바블리티의 차이점

1. 라이클리 후드는 확률이 아님 (확률은 다 합해서 1이여야 함. 라이클리 후드는 다 합해서 1인 아님)

\*우도는 함수 자체가 아니기 때문에 프로바블리티와 비교할 수 없음!

확률이면 1이지만, 1이라고 해서 무조건 확률은 아님!

* 나이브 베이시안의 약점

1.

* 시멘틱(문장안에 숨겨진 의도 찾기-예를 들어 목포 조건부 만남 이라는 문장에서 스팸인 것은 확인하는 것)의 경우, 베이시안을 사용할 수 없음.

2. 불충분한 양의 트레이닝 셋

한 단어가 안나왔다고 해서 0 분의 전체 확률이 되고 라이클리 후드를 다 곱하니까 분자의 값이 영이 되버림 -> 그래서 스무딩을 하는 것임.

* 라플라시안 스무딩

지시 함수(인디케이스 펑션)= 제이가 아이랑 같을때의 값을 구하는 것( 같을 떄의 확률)

제트가 2일 확률인 3/ 10 이 였는데, 이것은 4/ 10+케이로 바꿈.

* 1을 더해주는 이유는 ? 1)지금 관찰 되지 않는 값을 가질 수 있기 때문에 각각의 라이클리 후드에 일을 더하는 것임.2) 라이클리 후드값을 이용해서 확률을 구하는 것임! -> 라이클리 값에 1을 더하는 것이고 그리고 이건을 확률이니까 (확률값에 대한 라이클리 후드이니까) 확률의 성격(확률의 성격은 다 더해서 1임)을 유지하기 위해 분모에 케이(1을 더하는 변수의 수만큼, 코퍼스 개수의 값)를 더하는 것임.
* 사전확률(예시에서 23/102은 뭐임?)

사전확률의 추정치를 사용할 수 밖에 없음.

사전확률을 알수가 없으므로 곱을 0으로 만들지 않기 위해 사전확률에 1을 더 해줌.

\*베이즈 정리 외워!

**<강의 노트 12\_엠엘이\_엠에이피>**

- 맥시멈 라이클리후드 추정

데이터를 얻을 수 있는/ 관찰 할 수 있는 ㄱ밧

통상적으로 확률이라고 부르나 확률은 아니다(다 더해서 1아님)

세터1(동전이 던졌을 때 앞면이 나올 확률=1/3), 세터2(1/2), 세터3(1/4)

동전을 던질 결과가 10개 있는 경우(1000100111) -> 주어진 관찰 값

라이클리 후드는 나에게 주어진 관찰 값이 있을 때, 주어진 관찰 값을 통해 얻어낼 수 있는 함수 값( 모든 모수 별로 관찰값이 나올 확률을 확인 하는 거임/ 그중 가장 큰 확률의 함수 값이 맥시멈 라이클리 후드 추정임)

우도란 관찰값을 발생시키는 확률(수치) 임.

세터1,2,3에 따라 라이클리 후드값은 다 다름.(라이클리1,2,3 생성)

이때, 라이클리 후드값이 가장 큰 것응 찾을 것.

라이클리 후드는 현재 관찰값을 발생 시킬 함수 값(수치 값)을 의미함.

모수에 따라 함수 값이 다 달라질텐데, 그 중 함수 값이 가장 큰 것이 맥시멈 라이클리 후드 에스티메이트임.

라이클리 후드는 순서는 상관 없이 값만 신경 씀!

우도를 정의할때는 세터가 주어졌을 때, 관찰된 값을 얻을 수 있는 수치

베이시스 룰

관찰값이 (100011101)라고 주어졌을 때(이때 세터는 앞면이 나올 확률)

* 라이클리 후드(를 정리를 하면)는 세터의 함수가 나옴/ 라이클리 후드가 세터의 값에 따라 달라짐. 라이클리 후드는 세터의 함수이지 세터의 pdf(probablity density function)이 아님.

우도값을 구할 수 있기 때문에 사후 확률을 구할 수 있음.

< MLE vs MAP>

프라이어는 세터의 분포 값임

사후 확률은 사전 확률과 라이클리후드의 곱이니까 저런 모양의 그래프가 나옴.

우리가 원하는 것은 사후 확률임.

따라서 예시1와 같은 경우, 라이클리 후드만 쓰면 되고(MLE)

예시 2의 그림과 같은 경우 라이클리 후드만 쓰면 안됨!

우리가 최종적으로 구해야 하는 것은 MAP 임!

사전 확률을 정확하게 못구해도 추정치는 구해야 하며, 사후 확률을 구하기 위해서 엠엘이만 쓰기도 한다.

최대 우도 추정이란 관측 값이 주어졌을 때, 그 관측값이 가장크게 나올 확률!

라이클리 후드는 모수의 함수임!(우도의 정의, 각 관찰값들이 나올 확률을 싹 곱한 것)

우도를 수학적으로 최대한 하면 그것이 엠엘이 임.

볼록함수니까 미분해서 0이 되는 구간이 최대값 혹은 최소값이 될 수 있는 후보군이고

그것을 다시 한번 미분하면 최대갑 혹은 최소값을 구할 수 있음

엠엘이 예제는 면접 질문 될 수 있음.

우도 계산할 때 100프로 로그 우도를 씀 (왜? 증가 값이니까 로그를 사용함)

로그의 곱은 로그의 합이 되고

적어도 로그의 미분은 알아둘 것.

데이터 사이언스에서 로그는 밑이 자연 로그임!

2개도 함수 ( 3번 미분) 해서 양수면 최소값, 음수면 최대 값임(2개도 함수의 최소,최대)

엠엘이 예제는 손으로 풀수 있어야 함!

핸즈온 머신러님 빼고 머신 러닝 책 추천 안해줌.

* Pattern Recoginition and Machine Learning(Bishop): 1.5 Decision Theory (번역본있음)-> 머신러닝의 바이블 (현재 바이블)-> 여러 번 봤는지 박사과정에서 물어봤다고 함..
* Pattern Classification(Duda) : 2. Bayesian Decision Theory (번역본있음) (옛날 바이블)

나이브 베이시안을 전처리 작업으로 사용함.: 전체 문장을 잘라서 재 태깅 함.