워크북

교과목명 : 머신 러닝 차시명: 1차시

◈ 담당교수: **장 필 훈**

◉ 세부목차 (학습하기에서의 대단원, 중단원 명) • 기계학습의 개념 • 본 과정에서 다룰 범위 • 곡선피팅문제 • 베이즈 정리

	학습에 앞서

■ 학습개요

기계학습의 넓은 범위 중 패턴인식에 한정하여 학습한다. 데이터과학의 기초를 이루는 부분이며, 앞으로 더 깊은 공부를 위해서 필수적인 과정이다. 가장 대표적으로 곡선피팅문제를 소개하고 간단히 해결과정을 보인 후, 베이즈 정리를 이용해 결론을 도출해본다. 그리고 확률분포중 가장 자주 쓰이는 가우시안 분포를 살펴본다.

■ 학습목표

1	패턴인식문제가 어떤 것인지 이해한다
2	확률분포와 베이즈정리에 대한 기본적 이해를 가진다

■ 주요용어

용어	해설
비지도학습	지도학습은 문제와 답을 모두 주지만, 비지도학습은 문제만 주어진다. 예를 들어 군집화 문제는 데이터셋을 적절히 비슷한것들끼리 묶는 문제인데 이 경우는 답을 주지 않고 학습하기 때문에 비지도학습이라 부른다.
과적합	학습과정에서 오차가 0이 되도록 훈련집합에 대해 완벽하게 학습이 가능하지만, 새로운 데이터셋이 들어왔을 때 예측력은 현저하게 떨어지는 경우 모델이 과적합되었다고 한다. 모델이 추상화를 통해 문제를 '이해'한 것이 아니라 문제와 답을 외운것과 같은 현상.
확률분포	모집단 전체에 대해 샘플이 어떤식으로 모여있는지 알수있는 분포. 확률변수가 어떤값을 가질 수 있는지 나타내는 분포를 확률분포라고 하는데 이때 확률변수가 샘플 하나를 말한다. 예를들어 완전히 균일한 확률로 퍼져있다면 uniform distribution이라고 할 수 있다. 이번 수업에는 분포중 가장 많이 쓰이는 가우시안분포에 대해서 배운다.

학습하기

오늘은 기계학습의 전반적인 개념과 본 과정에서 그 중에 무엇을 배울지 이해하는 것이 목표입니다. 우리가 다루는 문제는 필수적으로 확률을 포함하고 있으므로 확률분포에 관해서도 배우겠습니다. 그 외에도 앞으로 학습을 하면서 배운 이론들을 실제 데이터에 적용해 보기 위해 필요한 툴에 관해 알아 보겠습니다.

기계 학습이라고 하면 넓게는 생각하는 기계를 만드는 모든 과정을 통칭합니다 최근에 화제가 되고 있는 자율 주행이나 번역기는 모두 기계학습의 결과물 입니다 하지만 여러분이 알고 있는 그런것들은 상당히 고수준의 응용 단계 에 속하고 먼저 그것들을 이해하기 위해서는 기본적인 학습의 기초에 관해 알아야 합니다. 본 과정은 딥러닝 과정을 포함하고 있지 않으므로 주로 고전적인 기계 학습 방법에 관해 다룹니다. 그리고 고전적인 기계학습 문제들은 수학에 바탕을 둔 것이 많아서 수식을 적지 않게 다룹니다 기본적인 것들 위주로만 해도 수식을 많이 다루게 되니 학습자들은 그점을 염두에 두어야 합니다

고전적인 기계 학습은 거의 전부 패턴인식 문제 라고 해도 과언이 아니고, 패턴 인식은 곧 추상화 과정을 말합니다 추상화는 우리가 얻을 수 있는 많은 데이터를 통해서 그 안에 보이지 않는 어떤 규칙이 있는지 알아 내는 과정이고, 기계 학습은 그 과정을 기계가 할 수 있게 만드는 것이 목표입니다.

우리 주변의 많은 것으로부터 패턴을 관찰할 수 있는데 가장 대표적으로는 손글씨를 들 수 있습니

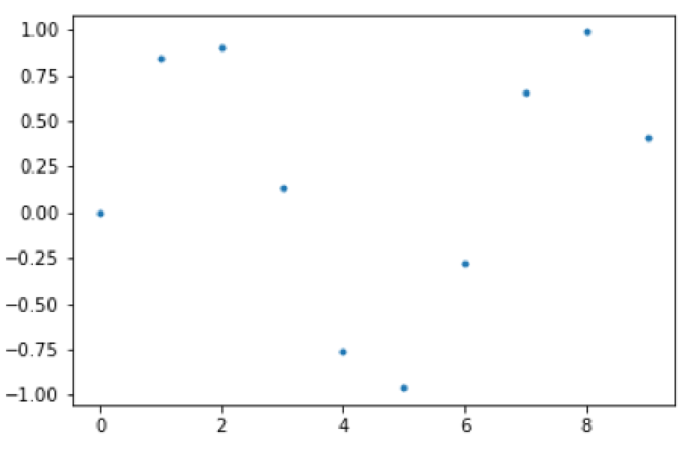
다. 사람의 손글씨는 사람마다 모두 조금씩 다르고 각자의 개성이 있지만 우리는 글씨를 있는데 큰 어려움이 없습니다 ㄱ의 모양은 천차만별이고 쓸때마다 달라지지만 한글을 배운 사람이라면 ㄱ(기역)을 읽는데 어려움을 느끼지 않습니다. 숫자를 예로 들자면 전 세계가 다 같은 모양의 아라비아숫자를 사용하지만 세상 모든 사람들이 숫자를 다른 모양으로 씁니다 단지 비슷할 뿐입니다 이렇게 비슷한 것들을 묶어서 같은 것으로 인식 할 수 있는 능력이 추상화 능력입니다 이미지 형태로 숫자를 읽어서 이것이 어떤 숫자인지 알아 맞추는 프로그램을 만드는 것은 언뜻 생각하면 쉬워 보이지만 모든 사람이 다양한 형태로 글씨를 쓰기 때문에 막상 해 보려면 쉽지 않습니다 패턴 인식 문제는 이런 문제를 해결하려는 노력의 결과물 입니다.

기계학습은 분류를 하자면 굉장히 다양한 형태로 분류할 수 있습니다 분류의 기준에 따라서 다양할 수 있지만 흔히 지도 학습 비 지도 학습 강화 학습 이렇게 세가지로 분류합니다 (우리의 목표는 패턴을 인식 하는 것이고 보통은 수학적인 함수($f(x)$)로 나타나는 간단한 문제들 이므로 앞으로는 수학적인 표현을 이용해서 문제를 표현하고 해결 하겠습니다) 지도 학습 의 경우 $y=f(x)$에서 x와 y가 모두 명시적으로 주어집니다. 그 다음 가장 '그럴듯'한 f 를 아는 것이 목표가 되는 것입니다. 비 지도 학습의 경우 y가 명시적으로 주어지지 않습니다. 대표적으로 클러스터링 즉 군집 화 문제를 들 수 있는데 많은 x들이 주어지면 우리는 비슷한 것들끼리 묶어볼 수 있습니다. 이 경우는 각각의 데이터 샘플이 어떤 클러스터에 속하는지 답이 처음부터 주어지는 것이 아니기 때문에 비 지도 학습으로 분류 됩니다 그리고 강화 학습을 들 수 있습니다. 강화 학습은 지도와 비지도 둘 어느쪽에도 넣지 않는데 그것은 강화 학습이 명확한 답을 처음부터 다 가지고 있다고 보기도 힘들고 그렇다고 가지고 있지 않다고 보기도 힘들기 때문에 따로 분류합니다.

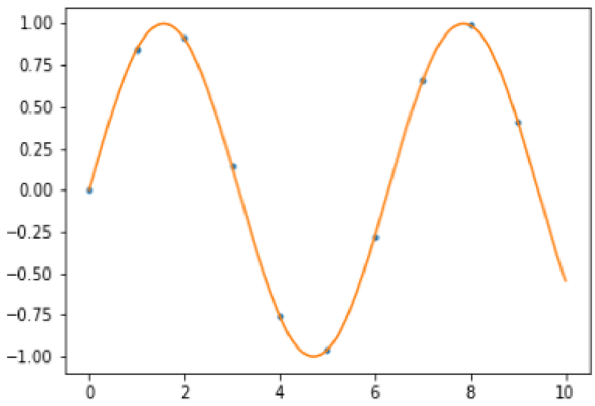
오늘은 패턴인식 문제의 가장 대표적인 문제인 다항식 곡선 피팅 문제를 보겠습니다.

<다항식 곡선피팅>

2차원 평면의 다음 점들을 보세요. 규칙을 알 수 있습니까?



위의 점들은 sin(x)의 곡선에서 임의로 열개를 뽑은 것입니다. 다음 그래프를 보세요.



패턴인식은, 위의 점만 찍힌 것들로부터 sin(x)를 어떻게 하면 얻어낼 수 있을까–하는 문제입니다.

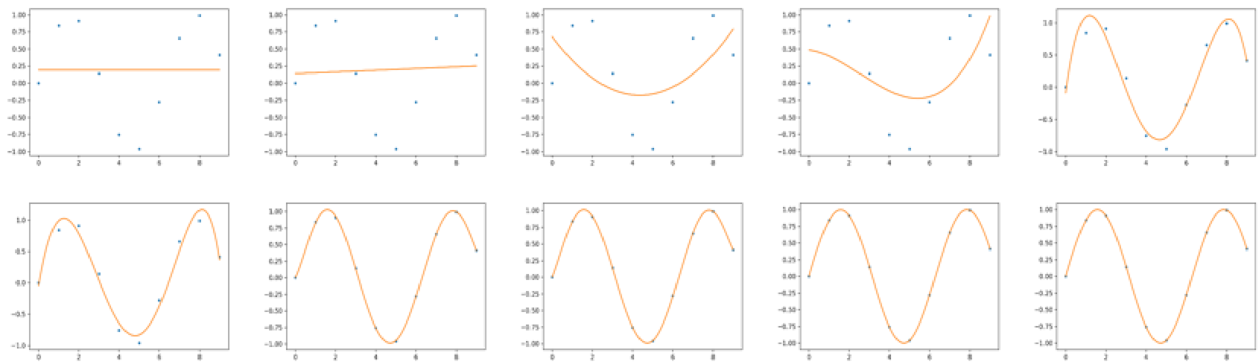
여기에는 가정이 필요합니다. ① 해당 점들이 어떤 곡선으로부터 얻어진 것이다 ② 그 곡선의 식은 다항식일 것이다. 이 두가지 가정입니다.

다항식은 아래 형식을 가진 식입니다. 그리고 이것은 우리의 '모델'입니다. 모델이란 데이터의 숨은 규칙을 설명해내는 추상화된 형식을 이릅니다.

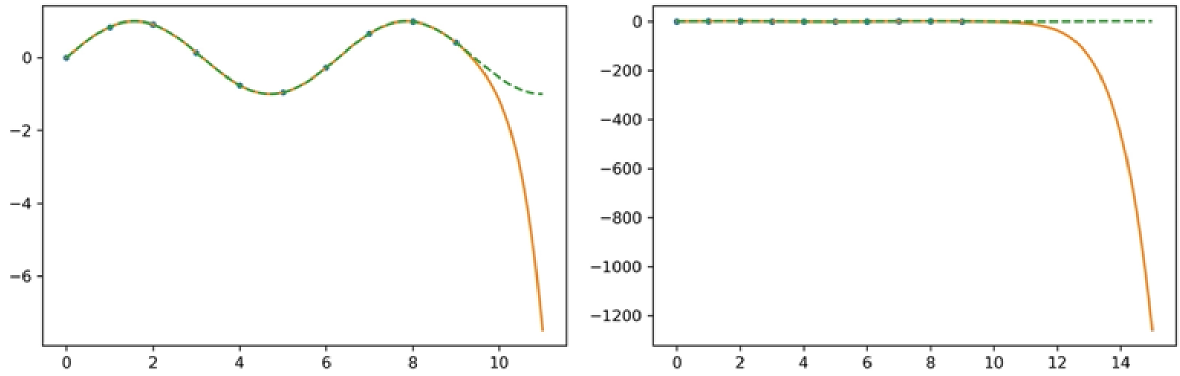
$$\sum_{j=0}^{M} w_j x^j = w + w_1 x + w_2 x^2 + \cdots + w_M x^M = y(x, w)$$

우리는 우선 이 다항식의 최고차항을 결정해야 합니다. 결정한 뒤, 주어진 점에 가장 잘 맞게 배열해볼 것입니다.

다음은 그 결과입니다. 0차부터 9차까지 보여줍니다.



대략 6, 7차부터 거의 완벽하게 fitting함을 알 수 있습니다. 그러면 답을 찾은 것일까요? 틀림없이 sin(x)와는 많이 다를 것입니다. 예를들어 x=11이나 12정도로 더 큰값이라면 어떨까요 다음을 보세요



x값이 커질수록 우리가 추정한 함수와 계속 더 멀어집니다.

이 예제에서 우리는 새로운 문제를 볼 수 있습니다.

① 어떻게 하면 데이터를 더 잘 예측하는 모델을 추정할 수 있을까

② 해당 모델이 정확함을 어떻게 하면 더 확신할 수 있을까?(모델평가의 문제)

1번문제는 뾰족한 수가 없습니다. 오랜 노하우와 넓은 배경지식이 도움이 됩니다. 2의 경우는 모델을 비교하는 척도를 만드는 문제입니다. 자주 쓰이는 척도들이 많이 있습니다. 예를들어 다음과 같은 함수는 오차의 제곱의 합을 구합니다.

$$\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2$$

다음과같은 함수는 어떨까요? 물론 가능합니다.

$$\sum_{n=1}^{N}|y(x_n, w) - t_n|^2$$

어떤 함수를 오차함수로 써야 할지는 상황에 따라 다르고 정확한 답이 없습니다. 다만, 자주 쓰이는 것들이 있으므로 특별한 경우가 아니면 우리도 관례를 따르면 크게 틀릴 염려가 없습니다.

실제 데이터는 필연적으로 잡음(노이즈)를 포함합니다. 자연에서 얻을 수 있는 데이터는 결코 완전할 수 없고, 그렇게 얻을 수 있다 해도 우리의 계산은 그 정확도에 한계가 있습니다. 그래서 오차는 본질적으로 0이 될 수 없다고 생각하면 됩니다. 그런 상황에서 어떻게 하면 오차들에도 불구하고 데이터 본연의 규칙을 찾아낼 수 있을까 연구하는 것이 패턴인식 분야입니다.

위에서 우리가 본 것은 '과적합'의 가장 대표적인 예입니다. 9차의 다항식이라면 x가 서로 다른 점 열개를 어떻게 잡아도 완벽하게 표현해낼 수 있습니다. 하지만 열한번째 점을 예측하는 것은 더 어려워집니다. 10개 점에 지나치게 알맞게 최적화되어 있기 때문입니다. 이런 현상을 막기 위해 가장 자주 쓰이는 강력한 방법은 정규화입니다. 차수가 높아지면서 데이터에 완벽하게 정합하려 하면 할수록 다항식의 w값이 커짐을 확인할 수 있는데, 여기에 착안해서 w의 크기도 손실함수에 집어넣는 것입니다.

$$\tilde{E}(w) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2}\|w\|^2$$

$$\|w\|^2 = w_0^2 + w_1^2 + w_2^2 + \cdots + w_M^2$$

그러면 왜 |w|가 아니고 ||w||^2일까요. 우리는 다음단원에서, 베이지언 확률을 이용해서 이것이 최적임을 보일 것입니다. 또, |w|를 사용했을 때, 혹은 다른 형태를 사용했을 때 특징에 관해 다음시간인 선형회귀에서 더 자세히 다룰 것입니다.

<확률>

앞으로도 많이 보게 될 베이즈 정리에 대해 배우겠습니다.

아래는 확률의 곱의 법칙입니다

$$p(X,Y) = p(Y|X)p(X)$$

아래는 확률의 합의 법칙입니다

$$p(X) = \sum_{Y} p(X,Y)$$

위 둘을 이용해서 다음을 얻을 수 있습니다.

$$p(X,Y) = p(Y,X) = p(Y|X)p(X) = p(X|Y)p(Y)$$

위 식의 오른쪽 두 항을 변형하면 다음을 얻습니다

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

여기서 왼쪽은 사후확률(posterior), 오른쪽의 분자는 likelihood(우도)와 사전확률(prior), 그리고 분모는 evidence라고 합니다. 보통 증거(evidence)는 고정되므로 다음과 같이 쓸 수 있습니다.

$$p(Y|X) \propto p(X|Y)p(Y)$$

여기서 베이지안 정리의 강력함을 볼 수 있습니다. 사전확률을 사후확률로, 사후확률을 사전확률로 바꿀 수 있습니다.

앞서 다항함수피팅문제에서 생각해보면, 우리가 이미 가지고 있는 데이터 열 개는 evidence에 해당합니다. 이 데이터 열 개가 특정한 분포에서 추출 되었다고 가정하고 어떤 분포를 설정할 때 이 데

이터 열 개를 얻을 확률이 가장 높은지 베이즈 정리를 이용해서 찾아 볼 수 있습니다. 우리는 다항 함수를 가정 했으므로 우리가 가정한 분포는 모두 다항 함수 형태를 가집니다. 따라서 W를 추정 하는 문제가 됩니다. W를 어떻게 잡으면 우리가 이미 가지고 있는 데이터 열 개를 받을 확률이 가장 높아 질 까요 다시 말하면 특정 분포에서 데이터 열 개를 뽑았을 때, 어떤 분포를 가정해야 그 데이터 열 개가 우리가 가진 데이터 10개 와 가장 비슷할까요. 이 문제를 최대 가능도 문제라고 합니다. 수식으로 적으면 p(D|w)를 최대화하는 문제 입니다.

데이터에는 필연적으로 노이즈가 포함되어 있고 노이즈는 가우시안 분포를 다룬다고 보는 것이 일반적이므로 먼저 확률분포 특히 가우시안 분포에 대해 더 자세히 알아 보겠습니다.

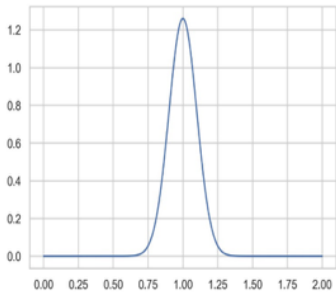
가우시안 분포는 식은 다음과 같습니다

$$N(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

더 알아보기 쉽게 적으면 다음과 같습니다 동일한 식입니다

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

이것을 그래프로 그려 보면 다음과 같은 모양이 됩니다 다음은 평균이 1이고 표준편차가 0.1일때 입니다



먼저 확률분포의 기본 적인 성질 하나를 확인해 보겠습니다. 확률분포는 정의역의 전 구간에 걸쳐 확률 밀도 함수를 적분했을때 1이 나와야 합니다. 식으로 쓰면 아래와 같습니다

$$\int_{-\infty}^{\infty} N(x|\mu,\sigma^2)dx = 1$$

본격적인 계산에 들어가기 전에 다음 식의 전개를 한번 보세요

$$\left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)^2 = \int_{-\infty}^{\infty} e^{-x^2}\,dx \int_{-\infty}^{\infty} e^{-y^2}\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy$$

그리고 구면좌표계에서 다음과 같이 쓸 수 있음도 기억해 두세요

$$e^{-(x^2+y^2)} = e^{-r^2}$$

위 둘을 이용해서 가우시안 적분을 한번 해 보겠습니다

$$\iint_{\mathcal{R}^2} e^{-(x^2+y^2)}dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\,drd\theta = 2\pi\int_0^{\infty} re^{-r^2}dr = 2\pi\int_{-\infty}^{0}\frac{1}{2}e^s ds$$

s=-r^2 이므로 ds=-2r dr이고, 다음도 이용했습니다.

$$\int_0^{\infty} re^{-r^2}dr = \int_0^{\infty}\frac{1}{2}e^{-r^2}\cdot 2r\,dr = \int_{\infty}^{0}\frac{1}{2}e^{-r^2}\cdot(-2r)\,dr$$

계속하면,

$$= \pi\int_{-\infty}^{0} e^{s}ds = \pi\left(e^{0} - e^{-\infty}\right) = \pi$$

결론적으로 우리는 가우시안 적분의 결과를 다음과 같이 얻을 수 있습니다

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2}\,dx = \sqrt{\frac{\pi}{a}}$$

위의 성질을 이용해서 가우시안 적분의 전체 구간에 대한 적분이 1을 가짐을 각자 시도해 보세요.

다음으로 가우시안 분포의 기대 값을 구해볼텐데 전체 구간적분보다 더 어려운 문제이므로 전체 적분값이 1을 가지는 자세한 과정은 생략하도록 하겠습니다

기댓값을 구해보겠습니다

$$\mathrm{E(x)} = \int_{-\infty}^{\infty} N(x|\mu,\sigma^2)x\,dx$$

정규분포식을 대입해서 전개해 보겠습니다

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\infty}^{\infty} x\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x$$
$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(\sqrt{2}\sigma t + \mu\right)\exp(-t^2)\,\mathrm{d}t$$
$$= \frac{1}{\sqrt{\pi}}\left(\sqrt{2}\sigma\int_{-\infty}^{\infty} t\exp(-t^2)\,\mathrm{d}t + \mu\int_{-\infty}^{\infty}\exp(-t^2)\,\mathrm{d}t\right)$$
$$= \frac{1}{\sqrt{\pi}}\left(\sqrt{2}\sigma\left[-\frac{1}{2}\exp(-t^2)\right]_{-\infty}^{\infty} + \mu\sqrt{\pi}\right) = \frac{\mu\sqrt{\pi}}{\sqrt{\pi}} = \mu$$

두 번째줄의 t는 다음과 같습니다 치환적분을 이용하는 것입니다

$$\mathrm{t} = \frac{\mathrm{x}-\mu}{\sqrt{2}\sigma}$$

결론적으로 우리는 기대값 μ를 얻습니다.

다음으로 분산을 구해보겠습니다. 분산을 구하기 위해 x제곱의 기대 값을 먼저 구해보겠습니다. 우선 다음 적분을 기억하세요

$$\int xe^{x^2}\,\mathrm{d}x = \int\frac{\mathrm{e}^t}{2}\mathrm{d}t = \frac{1}{2}\int \mathrm{e}^t\,\mathrm{d}t = \frac{1}{2}\mathrm{e}^t + C = \frac{1}{2}\mathrm{e}^{x^2} + C$$

$$t = x^2$$
$$dt = 2x\,dt$$

위를 이용해서 다음을 계산할 수 있습니다

$$E[x^2] = \int_{-\infty}^{\infty} N(x|\mu,\sigma^2)x^2 dx = \mu^2 + \sigma^2$$

X 제곱 의 기대값과 X 의 기대값을 이용해서 분산을 구하는 것은 고등학교 때 이미 배웠습니다.

가우시언분포의 기본적인 성질을 알아 보았습니다. 노이즈가 이 가우시안 분포를 따른다고 할 때 어떻게 하면 가장 좋은 w를 찾을 수 있을까요. 이것을 MLE문제라고 합니다. 앞에서 설명한 것과 같이 우도(likelihood)를 최대화하는 문제입니다.

그 문제를 풀어보겠습니다. 만약 가우시안 분포에서 무작위로 데이터를 추출 한다면 조건부 확률은 다음과 같습니다 우리는 데이터를 추출 할 때 다른 데이터들의 영향을 받지 않으므로 likelihood를 모두 곱하는 것과 같습니다.

$$p(X|\mu,\sigma^2) = \prod_{n=1}^{N} N(x_n|\mu,\sigma^2)$$

$$\boldsymbol{X = x_1, x_2, \ldots, x_N}$$

여기서 이것이 확률의 곱이 아니라는 점을 잊으면 안 됩니다. 우도와 확률은 다릅니다.

동전 던지기를 예로 들면, 동전의 앞면이과 뒷면이 나올 확률이 같다고 가정할 때, 동전을 두번 던졌을 때 앞면이 두번 나올 확률은 1/2*1/2=0.25와 같습니다. 이것을 가능도 함수에 관해 쓰면 다음과 같습니다

$$L(P_H = 0.5\,|HH) = 0.25$$

즉 '2번 앞면을 관찰했을 때 동전의 앞면이 나올 확률이 0.5인 likelihood는 0.25라고 말합니다. 여기서는 확률 자체가 변수가 되는 것입니다. 우도는 확률처럼 답 한다고 1이 나올 필요가 없습니다. 말 그대로 어떤 변수가 얼마나 더 잘 나올 만한 것인지 그 가능도를 보는 것이기 때문에 상대적인 비교만 가능할뿐 1이 되지 않거나 넘어도 됩니다.

자 이제 주어진 X 값에 대한 t값 즉 목표값이 y(x,w) 를 평균으로 하는 가우시안 분포를 따른다고 합시다. 노이즈가 가우시안 분포를 따른다는 말과 같은 말입니다. 그러면 다음과 같이 적을 수 있습니다

$$p(t|x,w,\beta) = \mathrm{N}(\mathrm{t}|\mathrm{y}(\mathrm{x},\mathrm{w}),\beta^{-1}), \qquad \beta = \frac{1}{\sigma^2}$$

$$p(t|\boldsymbol{x},w,\beta) = \prod_{n=1}^{N} N(t_n|y(x_n,w),\beta^{-1})$$

이 식을 최대화하는 W를 구할 텐데 우선 로그를 취하겠습니다. 로그함수는 2가지 장점이 있습니다 첫째로 로그함수는 단조 증가함수이므로 로그를 취한뒤 최댓값을 구하면 원래 함수의 최댓값을 구하는 것과 같습니다. 그리고 둘째로 로그를 취하면 곱이 합으로 바뀌므로 컴퓨터로 계산할 때 수치적으로 더 stable합니다. 여러 가지 숫자를 곱하는 데 중간에 너무 작은 숫자나 너무 큰 숫자가 들어가면 컴퓨터는 정확도의 한계로 인해 정답과 아주 먼 결과 값을 주기도 합니다. 하지만 덧셈의 경우는

너무 작은 값이나 큰 값이 결과 값의 오차를 크게 바꾸지 못하므로 이론을 넘어서 실제 적용할 때 더 좋은 결과를 보여 줍니다

위 식에 로그를 취해 봅시다

$$\ln \prod_{n=1}^{N} N(t_n | y(x_n, w), \beta^{-1})$$
$$= \ln \prod_{n=1}^{N} \frac{\beta}{\sqrt{2\pi}} \exp\left\{-\frac{\beta^2}{2}\left(t - y(x, w)\right)^2\right\}$$
$$= -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi)$$

W와 관계된 항은 첫 번째 항 뿐입니다. 따라서 다음 값을 최소화 하는것과 같습니다.

$$\sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2$$

우리는 앞서 오차 함수를 디자인할 때 제곱 함수를 썼습니다. 그리고 지금 베이즈 정리를 이용해서 노이즈가 가우시안 분포를 따를 때 이 제곱합 오차 함수가 W를 추측하는데 가장 좋은 방법 이라는 것을 보였습니다.

	연습문제

1. (OX문제) 클러스터링이나 밀도 추정 문제는 지도 학습에 속한다

X 데이터셋 하나에 목표값 t하나가 주어지는 것이 아니므로 비 지도 학습에 속한다

2. (OX문제) 베이즈정리를 이용하면 사전 확률을 사후 확률을 이용해서 나타낼 수 있다

O 사후확률은 사전확률*가능도에 비례한다

3. (OX문제) 가우시안 분포는 정규화 되어 있다

O 확률 변수의 전 영역에 걸쳐 적분한 결과가 1이면 정규화 되어 있다고 하고, 가우시안 분포는 그에 해당한다.

4. (OX문제) 우도(가능도, likelihood)의 합은 1이다.

X 가능도는 말 그대로 어떤 일이 얼마만큼 있음직 하냐를 나타내는 것이다. 그래서 1이 넘을수도 있고 안될수도 있다.

정리하기

1. 고전적인 기계학습은 패턴인식을 말하며, 학습방법에 따라 지도학습, 비지도학습, 강화학습으로 나뉜다.
2. 다항식 곡선피팅은 다음과 같은 식을 데이터에 맞추어보는 과정이다.

$$\sum_{j=0}^{M} w_j x^j = w + w_1 x + w_2 x^2 + \cdots + w_M x^M = y(x, w)$$

3. 다음을 베이즈정리라고 한다. 사후확률과 사전확률간의 관계를 알 수 있다.

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

4. 다음을 정규분포(가우시안분포)라고 한다.

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

참고하기

- Bishop, C. M. "Bishop-Pattern Recognition and Machine Learning-Springer 2006." Antimicrob. Agents Chemother (2014): 03728-14.

다음 차시 예고

다음시간에는 기저(basis)함수모델과 편향분산분해에 관해 배우겠습니다.