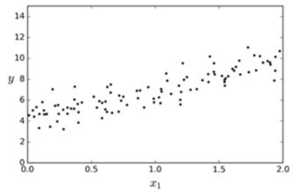
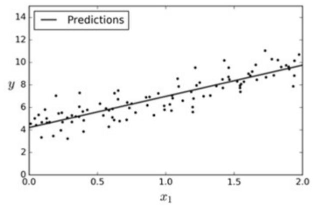
1.1 회귀 분석

- 데이터들 사이의 상관관계 또는 추이를 예측하거나, 대상 값 자체를 예측하는 지도학습 알고리즘
- 예를 들어, 방의 개수와 집값의 상관 관계 또는 과거 10년간의 영업 실적을 분석하여 미래의 영업 실적을 예측하는 것
- **하나의 종속변수와 하나 이상의 독립변수 간의 상관관계를 기본으로 하여 하나의 n차 선형방정식으로 변수 관계를 일반화하는 분석방법**

1.1 회귀 분석

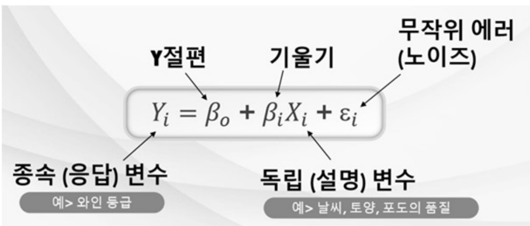
- **선형성(Linearity)**
  두 변수의 관계가 하나의 직선의 형태로 설명될 수 있는 관계를 지닌다는 것





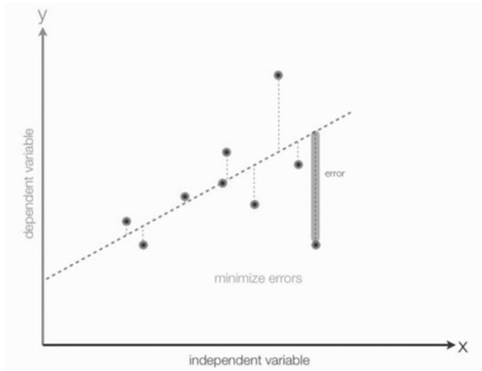
1.1 회귀 분석

- **선형 방정식 :** 두 변수의 관계를 방정식으로 표현
- **기울기**는 가중치의 역할을 하기 때문에 이걸 통해 x가 y에 얼마나 영향을 주는지 그 크기와 방향을 알 수 있음
- **절편**은 y = bx 라는 회귀선을 얼마나 위 또는 아래로 평행이동 시키는지를 정해줌



1.1 회귀 분석

- **선형 방정식**
  두 변수의 관계를 방정식으로 표현



1.2 최소제곱법Ordinary Least Square

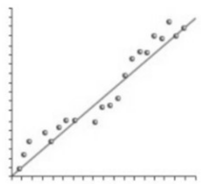
- 1781년에 허셸(W. Hershel)은 자신이 제작한 망원경을 사용하여 천왕성을 발견
- 1801년에 천문학자 피아치(J. Piazzi)는 이탈리아 팔레르모 천문대에서 소행성 하나를 관측했고, 이것을 세레스(Ceres)라고 명명하고 41일 동안 22개의 관찰 자료를 만듦
- 그러나, 얼마 후 세레스는 시야에서 사라져서 더 이상 관찰할 수 없게 되자, 당시 과학자들은 세레스의 궤도를 계산하여 출현 위치를 먼저 알아내기 위해서 서로 경쟁하게 됨

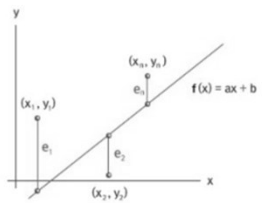
1.2 최소제곱법Ordinary Least Square

- 불과 18살이었던 가우스(C.F. Gauss)는 세레스의 궤도가 원뿔곡선이라는 가정 아래 '최소제곱'이라는 새로운 방법을 사용하여
- 피아치가 남긴 22개의 관찰 자료에만 의존해서 가우스가 예측한 위치에 11월 25일과 12월 31일 사이에 세레스를 재발견할 수 있었음
- 가우스는 최소제곱법을 사용하여 새로운 행성이 발견되는 대로 그 궤도를 계산해 낼 수 있었다.

1.2 최소제곱법

- 최적의 회귀모델은 실제 값 Y 와 예측 값 y의 차이가 최소가 되어야 함
- 데이터 분포와 가장 일치하는 회귀 모델을 구하는 방법 중 하나는 **최소제곱법(Ordinary Least Square; OLS)**을 이용
- 즉, 오차(실제값과 예측값 사이 차이)의 값은 무조건 양수이어야 함
  => 따라서, 이 값을 제곱시킴





1.2 최소제곱법

- 최적의 최소제곱 직선

$$E(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

- 미적분학의 이론에 의한 편미분

$$\begin{cases} 0 = \dfrac{\partial E}{\partial a} = \sum 2(y_i - ax_i - b)(-x_i) = 2(a\sum x_i^2 + b\sum x_i - \sum x_i y_i) \\ 0 = \dfrac{\partial E}{\partial b} = \sum 2(y_i - ax_i - b)(-1) = 2(a\sum x_i + b\sum 1 - \sum y_i) \end{cases}$$

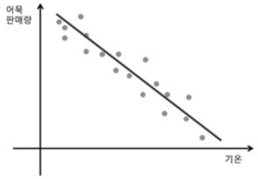
- 연립방정식

$$\begin{cases} a\sum x_i^2 + b\sum x_i = \sum x_i y_i \\ a\sum x_i + b\sum 1 = \sum y_i \end{cases}$$

1.3 평균제곱오차MSE

- 비용함수

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



- 편미분

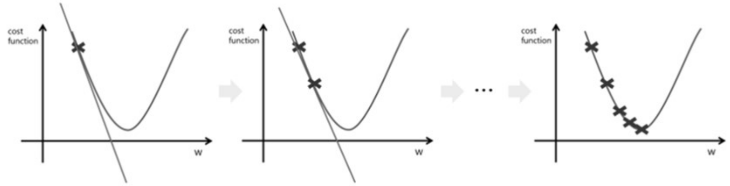
$$\begin{aligned} MSE &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \\ &= \frac{1}{n}\{(y_1 - wx_1 - b)^2 + \cdots + (y_n - wx_n - b)^2\} \\ &= \frac{1}{n}\{(x_1^2 + \cdots + x_n^2)w^2 + \bigstar\} \qquad (for\ \ w) \\ &= \frac{1}{n}\{nb^2 + \blacksquare\} \qquad (for\ \ b) \end{aligned}$$

1.3 평균제곱오차MSE

- 학습률에 의한 반복법

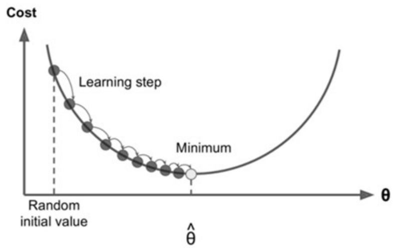
$$w \ := \ w \ - \ \alpha\frac{\partial}{\partial w}MSE \qquad b \ := \ b \ - \ \alpha\frac{\partial}{\partial b}MSE$$

$$x_0 = x_0 - \eta\frac{\partial f}{\partial x_0} \qquad x_1 = x_1 - \eta\frac{\partial f}{\partial x_1}$$



1.4 경사하강법Gradient Descent

- 손실 함수loss function 또는 비용 함수cost function라 불리는 목적 함수를 정의하고 이 함수의 값이 최소화되는 매개변수를 찾는 방법
- 손실 함수의 기울기를 이용해 손실 함수의 극소값을 찾는 알고리즘
- 함수의 최소값을 찾기 위해 현재 위치(보통 임의의 위치)에서 시작해서 기울기를 적절한 간격으로 조금씩 낮춰가며 적절한 최저점의 매개변수를 찾아냄

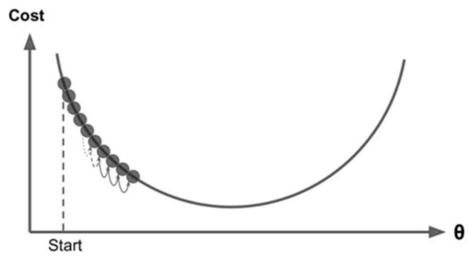


1.4 경사하강법Gradient Descent

- 지나치게 작은 학습률learning rate

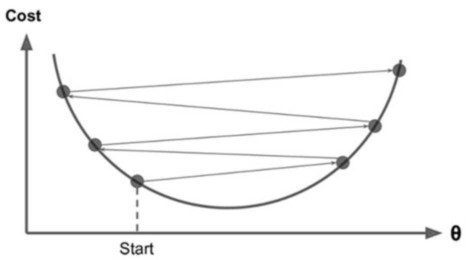
$$x_0 = x_0 - \eta\frac{\partial f}{\partial x_0}$$

$$x_1 = x_1 - \eta\frac{\partial f}{\partial x_1}$$



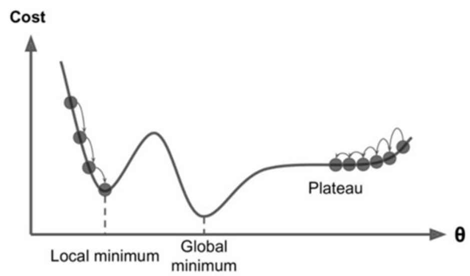
1.4 경사하강법Gradient Descent

- 지나치게 큰 학습률learning rate



1.4 경사하강법Gradient Descent

- 최적점이 여러 개 존재하는 경우의 학습률learning rate

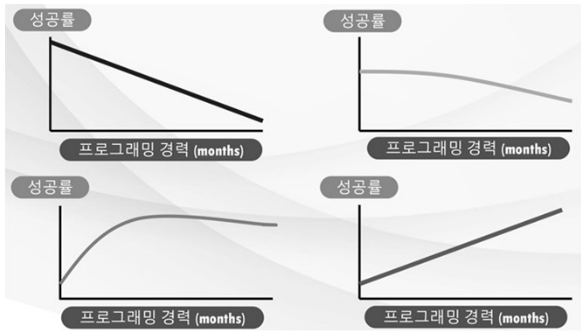


1.5 hyper parameter

- **학습률 같은 매개변수를 초매개변수라고 함**
- 가중치와 편향 같은 신경망의 매개변수와는 성질이 다른 매개변수
- 신경망의 가중치 매개변수는 훈련 데이터와 학습 알고리즘에 의해서 자동으로 획득되는 매개변수인 반면, **학습률 같은 초매개변수는 사람이 직접 설정해야 하는 매개변수**
- 일반적으로 이들 **초매개변수**는 여러 후보 값 중에서 시험을 통해 가장 잘 학습하는 값을 찾는 과정을 거쳐야 함
- 사람이 결정해야하는 것이기에 값을 결정하기까지 많은 시행착오를 필요로 함

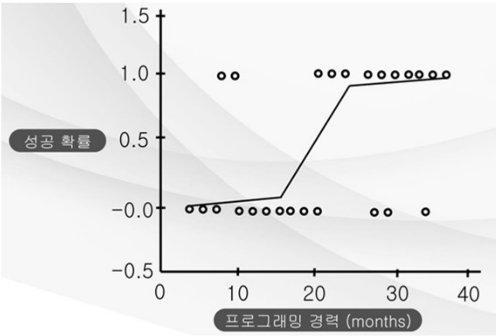
1.6 선형성으로 풀리지 않는 문제

- 프로그래밍 경력에 따른 자격증 시험 합격률



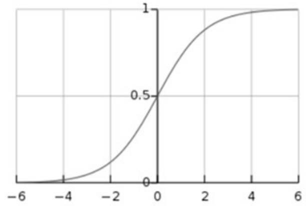
1.6 선형성으로 풀리지 않는 문제

- 프로그래밍 경력에 따른 자격증 시험 합격률



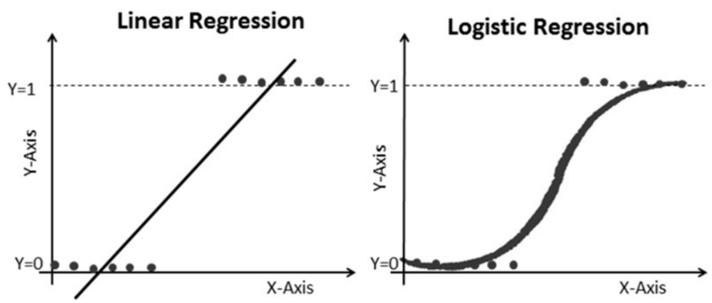
2.1 로지스틱 회귀

- 일반적인 선형 모델의 특수한 경우로 간주 => 선형 회귀와 유사
- 로지스틱 회귀는 선형 회귀와는 다르게 종속 변수가 **범주형 데이터**
- 해당 데이터의 결과가 범주로 나눠지기 때문에 일종의 분류에 속함
- 로지스틱 회귀는 종속 변수에 따라 binomial, multinomial 등으로 나눔



$$f(x) = \frac{1}{1 + e^{-x}}$$

2.1 로지스틱 회귀



2.2 로짓함수

- 이항 로지스틱 회귀의 모델의 경우 종속변수 y의 범위는 [0,1]로 제한됨
- 종속 변수가 이항이므로 정규분포 대신 이항분포를 따름

- 0부터 1까지로 값이 제한된 확률 값의 범위를
  0부터 ∞로 확장하기 위해 odds 비를 사용함

- 0부터 ∞까지로 확장된 결과를 다시 -∞에서 ∞로 확장하기 위해
  odds 비의 식에 자연로그를 취함 => logit 변환

- 이제 양쪽의 식이 모두 -∞에서 ∞까지의 값을 갖게 되므로
  선형회귀식에 이것을 적용하면 로지스틱 함수가 유도됨

2.2 로짓함수

- 선형 회귀 방정식

$$y = \beta_0 + \beta X$$

- y를 odds ratio(성공확률/실패확률)로 간주함

$$\text{odds ratio} = \frac{p(y=1|x)}{1-p(y=1|x)}$$

- 이 확률값은 계수들에 대해 비선형(0,1)이기 때문에 선형으로 변환하기 위하여 자연로그를 취함 (y범위 : -∞ ~ +∞)

$$\mathrm{logit}(p) = \log\frac{p}{1-p} \qquad \ln\frac{p_i}{1-p_i} = \beta \cdot X_i$$

2.2 로짓함수

- 로짓함수를 선형 회귀 방정식에 적용시킴

$$\ln\frac{p_i}{1-p_i} = \beta \cdot X_i$$

- 관심 대상은 성공확률이므로 성공 확률 P에 대하여 이항

$$p_i = \mathrm{logit}^{-1}(\boldsymbol{\beta} \cdot \mathbf{X}_i) = \frac{1}{1+e^{-\boldsymbol{\beta} \cdot \mathbf{X}_i}}$$

2.3 odds

- odds는 Probability의 또 다른 표현법
- 어떠한 사건이 일어날 확률을 p라고 할 때 그 사건에 대한 odds는 **p / (1−p)** 로 구할 수 있음

- 흔히 도박 배팅이나 혹은 질병의 발병 확률을 표현할 때 많이 사용
- 예를 들어 경마에서 특정 말이 이길 확률이 75%라고 할 때 odds는 0.75 / 0.25가 되며 보다 통상적인 표현으로는 정수표현을 사용해서 3:1 이라고 표현 => **4게임 중 3게임은 이기고 1게임은 진다**

2.3 odds

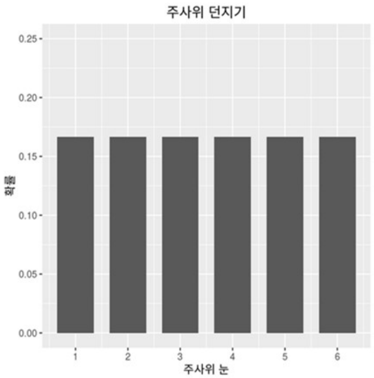
- 또한 Odds의 비(ratio), 즉 Odds ratio(승산비, 오즈비)을 이용하면 두 개의 property에 대한 연관성을 확인할 수 있음
- 예를 들어 특정 유전자(A)를 갖은 사람의 특정 질병(B)의 Odds ratio를 구하면 특정 유전자(A)와 특정 질병(B)의 연관성을 평가할 수 있음

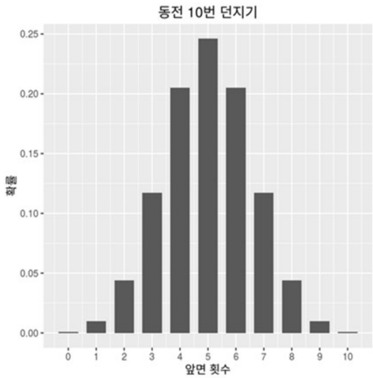
예시

		비만		
		Yes	No	Total
혈중 콜레스테롤	Normal	402	3614	4016
	High	101	345	446
		503	3959	4462

2.4 maximum likelihood

- 주사위의 각 눈이 나올 확률
- 동전을 10번 던져 앞면이 나올 확률



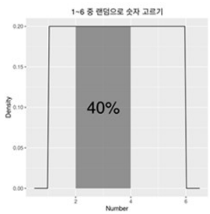


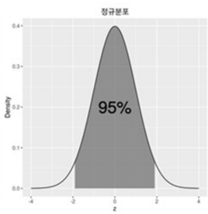
2.4 maximum likelihood

- 1에서 6 사이의 숫자 중 임의로 아무 숫자나 뽑을 확률을 구해 봄
- 정확히 5가 뽑힐 확률은 얼마일까?
- 1과 6사이에는 무한개의 숫자가 있으니 정확히 5가 뽑힐 확률은 1 / ∞ = 0
- 즉, 어떤 특정 숫자가 뽑힐 확률은 전부 0
- 이는 연속된 숫자 사이에서 뽑을 수 있는 숫자의 갯수는 무한하기 때문
- 따라서, 이런 연속사건인 경우 특정 숫자가 나올 확률을 말하는 것은 의미가 없음
- 대신, 숫자가 특정 구간에 속할 확률을 알아보는 것이 나음

2.4 maximum likelihood

- 1에서 6사이의 숫자 중 정확히 5가 뽑힐 확률은 0 이지만
- 4에서 5사이의 숫자가 뽑힐 확률은 20%임
- 전체 구간의 길이는 6-1=5이고 4에서 5사이 구간의 길이는 1이기 때문
- 마찬가지의 논리로 2에서 4사이의 숫자가 뽑힐 확률은 2/5=40
- 이처럼 우리는 특정 사건에 대한 확률 대신 특정 구간에 속할 확률을 구함으로서 간접적으로 특정 사건
- 이것을 설명하는 곡선이 **확률밀도함수(Probability Density Function)**





2.4 maximum likelihood

- A 주머니에 100원짜리 동전 2개와 10원짜리 동전 2개가 들어 있고, B 주머니에 100원짜리 동전 3개와 10원짜리 동전 2개가 들어 있다

- A, B 두 개의 주머니를 임의로 택한 후 2개의 동전을 꺼내보니 2개 모두 100원짜리 였다고 할 때 어떤 주머니에서 이 동전들이 나왔을 **가능성이 클까?**

2.4 maximum likelihood

- 정답은 B 주머니이다!
- 그 이유는 A 주머니로부터 2개의 100원짜리 동전을 꺼낼 확률은 **1/6**이고 B 주머니로부터 꺼낼 확률은 **3/10**이기 때문이다
- 선택한 주머니를 θ로, i번째 꺼낸 동전의 값을 X_i라고 할 때 이를 수학적 기호로 표시하면 X_1 = 100, X_2 = 100으로 관찰될 확률은 f(100, 100|θ)이다
- 다시 말해, 확률값을 최대로 하는 θ는 B 주머니라는 의미이다

$$L(\theta; x_1, x_2, \dots, x_n) = L(\theta; X) = f(X|\theta) = f(x_1, x_2, \dots, x_n|\theta)$$

$$\hat{\theta} = \arg\max_{\theta} L(\theta; X) = \arg\max_{\theta} f(X|\theta)$$

2.4 maximum likelihood

- 다시 말하면 f(100, 100|θ)를 최대로 하는 θ가 B 주머니라는 것은, 관찰된 2개 동전이 모두 100원짜리 일 때 B 주머니로부터 나왔을 가능성이 가장 높다(maximum likelihood) 것이다.
- 이러한 의미에서 주머니 B를 θ의 최우추정치(maximum likelihood estimation)라고 하고 θ를 구하는 방법을 최대우도추정법(the method of maximum likelihood estimation)이라 한다.

3.1.1 인공신경망과 생물신경망

- 사람의 뉴런
  - 두뇌의 가장 작은 정보처리 단위
  - 세포체는[cell body] 간단한 연산, 수상돌기는[dendrite] 신호 수신, 축삭은[axon] 처리 결과를 전송
  - 사람은 10^{11}개 정도의 뉴런을 가지며, 뉴런은 1000개 가량 다른 뉴런과 연결되어 있어 10^{14}개 정도의 연결

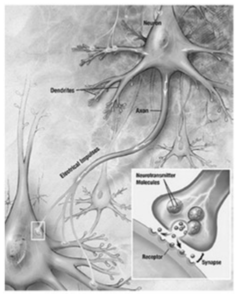


그림 3-1 **사람의 뉴런의 구조와 동작**