

- 1 Lab. 회귀분석과 최적화 문제
- 2 -Microsoft Excel을 이용하여 Machine Learning을 학습해 보자.
- 3
- 4 1. 준비물
- 5 1)Lab. 회귀분석과 최적화문제.xlsx
- 6
- 7 2. 7개의 Excel 함수 설명
- 8 1)SUM
- 9 -Cell 범위 수치들의 합을 계산한다.
- 10 -목적함수의 계산
- 11 2)SUMPRODUCT
- 12 -두 개의 지정된 범위에 있는 수치들을 서로 곱하고 그 합을 계산한다.
- 13 -입력이 선형합
- 14 3)SUMXMY2
- 15 -두 개의 지정된 범위에 있는 수치들의 차를 제공하고 그 합을 계산한다.
- 16 -제공오차의 산출
- 17 4)EXP
- 18 -지수함수의 값을 계산한다.
- 19 -Sigmoid 함수
- 20 5)MAX
- 21 -지정된 범위의 최대값을 구한다.
- 22 -MAX 풀링, ReLU
- 23 6)RAND
- 24 -0 이상 1 이하의 난수를 발생한다.
- 25 -초기값 설정
- 26 7)IF
- 27 -대소를 판정한다.
- 28 -Image의 판정
- 29
- 30
- 31 Task 1. SUMPRODUCT 함수 사용하기
- 32 1. 목적
- 33 -두 개의 지정된 범위에 있는 수치들을 서로 곱하고 그 합을 계산한다.
- 34
- 35 2. 준비물
- 36 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task1] tab
- 37
- 38 3. 주어진 문제
- 39 1)(x, y) = (0.9, 0.1), (a, b) = (0.8, 0.3) 일 때, 이 때 다음과 같이 수치들의 곱의 합 S를  
SUMPRODUCT 함수를 이용하여 구하시오.  
 $S = ax + by$
- 40
- 41
- 42 4. 해결
- 43 1)현재 B3 cell에 설정되어 있는 것처럼 =SUMPRODUCT(B1:B2,D1:D2) 로 구할 수 있다.
- 44 2)이와 같이 두 개의 지정된 범위에 있는 수치들을 서로 곱하고 그 합을 계산할 때에는 SUMPRODUCT를 이용한  
다.
- 45
- 46
- 47 Task 2. SUMXMY2 함수 사용하기
- 48 1. 목적
- 49 -두 개의 지정된 범위에 있는 수치들의 차를 제공하고 그 합을 계산한다.

50  
51 2. 준비물  
52 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task2] tab  
53  
54 3. 주어진 문제  
55 1)(x, y) = (0.9, 0.1), (a, b) = (0.8, 0.3) 일 때, 이 때 다음과 같이 수치들의 차의 제곱합 Q를  
SUMXMY2 함수를 이용하여 구하시오.  
56  $Q = (x - a)^2 + (y - b)^2$   
57  
58 4. 해결  
59 1)현재 B3 cell에 설정되어 있는 것처럼 =SUMXMY2(B1:B2,D1:D2) 로 구할 수 있다.  
60 2)이와 같이 두 개의 지정된 범위에 있는 수치들의 차를 제곱하고 그 합을 계산할 때에는 SUMXMY2를 이용한다.  
61  
62  
63 Task 3. EXP 함수 사용하기  
64 1. 목적  
65 -EXP 함수를 이용하여 Sigmoid 함수를 작성해 본다.  
66  
67 2. 준비물  
68 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task3] tab  
69  
70 3. 주어진 문제  
71 1)Sigmoid함수  $\sigma(x)$ 는 아래 기술하는 것처럼, 지수함수  $e^x$ 부터 다음과 같이 얻어진다.  
72 2) $\sigma(x) = 1 / 1 + e^{-x}$   
73 3)그래프는 sigmoid.png 그림 참조  
74 4)여기에서 e는 네피어 수라고 부르는 것으로 다음과 같은 근사값을 가진다.  
75  $e \approx 2.71828$   
76  
77 4. 해결  
78 1)EXP함수는 지수함수  $e^x$ 의 값을 구하기 위한 함수이다.  
79 2)위 3.2)에 따라 sigmoid 함수는 B2 cell에 설정되어 있는 것처럼 =1/(1+EXP(-A2))로 구할 수 있다.  
80  
81  
82 Task 4. Excel의 상대 참조 사용하기  
83 1. 목적  
84 -Spreadsheet의 표준적인 cell 참조는 상대참조이다.  
85 -어느 cell에 작성된 계산식을 다른 cell로 복사하면, 그 상대적인 이동만큼 계산식의 주소가 갱신된다.  
86  
87 2. 준비물  
88 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task4] tab  
89  
90 3. 주어진 문제  
91 1)다음과 같이 사원 3명의 신장과 체중 기록이 있다.  
92 2)이 사람들의 체질량 지수를 구해보자.  
93  
94 4. 해결  
95 1)먼저 첫 번째 사원의 BMI 계산식(=C2/(B2^2))을 cell D2에 입력한다.  
96 2)다음으로 cell D2에 입력한 계산식을 D3:D4에 복사한다.  
97 3)상대 참조는 자동적으로 참조 주소를 갱신한다.  
98  
99

100 Task 5. Excel의 절대 참조 사용하기

101 1. 목적

- 102 -어느 cell에 작성된 계산식을 다른 cell로 복사할때, 계산식의 주소를 갱신하지 않는 참조법이 절대참조이다.
- 103 -참조하는 cell 주소에 \$를 붙인다.

104

105 2. 준비물

- 106 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task5] tab

107

108 3. 주어진 문제

- 109 1)자산관리회사가 관리하는 3명의 고객 A, B, C의 자산액을 달러로 환산해 본다.

110

111 4. 해결

- 112 1)처음에는 고객 A의 달러 자산 계산식(=B3/\$B\$1)을 입력한다.
- 113 2)달러 / 원 비율이 \$b\$1이라고 cell 주소를 절대참조하는 것에 유의하자.
- 114 3)다음으로 cell C3를 C4:C5에 복사한다.
- 115 4)그러면 달러/원 비율의 cell 주소는 고정된다.

116

117

118 Task 6. Excel의 복합 참조 사용하기

119 1. 목적

- 120 -식의 표현에 상대참조와 절대참조를 섞은 참조법이 복합참조이다.
- 121 -참조하는 cell의 행 또는 열의 주소 한 쪽에 \$를 붙인다.
- 122 -이 참조법을 이용하여 '구구단 계산표'를 작성해 본다.

123

124 2. 준비물

- 125 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task6] tab

126

127 3. 주어진 문제

- 128 1)표의 위쪽과 왼쪽 숫자들의 곱을 구하여 '구구단 계산표'를 완성해 보자.

129

130 4. 해결

- 131 1)처음에는 1 x 1을 작성한다.
- 132 2)즉, B2 cell에 계산식 =\$A2\*B\$1를 넣는다.
- 133 3)다음에 이 cell의 함수를 표 전체에 복사한다.
- 134 4)\$를 붙이지 않은 행 또는 열의 주소가 갱신되고, \$를 붙인 행 또는 열의 주소는 갱신되지 않는 것을 알 수있다.

135

136

137 Task 7. Excel의 해 찾기 사용 방법

138 1. 목적

- 139 -Data 분석을 위해 작성한 수학 model에 포함된 매개변수를 결정하는 데 매우 편리한 도구가 Excel에 포함된 [해 찾기]이다.
- 140 -예제를 이용하여, Excel 해 찾기 이용법을 알아보자.

141

142 2. 준비물

- 143 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task7] tab

144

145 3. 주어진 문제

- 146 1)함수  $y = 3x^2 + 1$ 의 최솟값을 구하는 문제를 Excel의 해 찾기를 이용해 풀어본다.

147

148 4. 해결

- 149 1)주어진 x에 대한 함수 y의 식(=3\*B3^2+1)을 입력한다.

- 2)그리고 cell B3에 적당한 x의 값을 설정한다(현재 C3에 5를 입력했지만, 특별한 의미가 있는 것은 아니다).
- 3)해 찾기를 실행하기 위해, [데이터] 메뉴에 있는 [해 찾기] 항목을 선택한다.
- 4)만일 [해 찾기]가 설치되지 않았으면, 다음과 같이 실행한다.
  - [파일] tab의 [옵션] 메뉴를 클릭한다.
  - [Excel 옵션] 상자가 열리면 왼쪽 창 중에서 [추가 기능]을 선택한다.
  - 추가로 얻어진 box 중에서 아래에 있는 [관리] dropdown 메뉴 중 [Excel 추가 기능]을 선택하고, [이동] button을 click한다.
  - [추가 기능] 상자가 열리면 [해 찾기 추가 기능]에 check하고 [확인] button을 click한다.
  - 설치 작업이 진행된다.
  - 바르게 설치되면 [데이터] 메뉴에 [해 찾기] 항목을 확인할 수 있다.
- 5)[해 찾기 매개변수] 창이 나타나면, 다음과 같이 설정한다.
  - 목표 설정 : \$C\$3
  - 대상 : 최소
  - 변수 셀 변경 : \$B\$3
  - [제한되지 않는 변수를 음이 아닌 수로 설정] check
- 6)[해 찾기] button을 click하여 실행한다.
- 7)해 찾기가 최소값을 구하면 [해 찾기 결과] 상자가 나타난다.
  - [해 찾기 해 보존] check 되어 있음
  - 메시지 상자에 "해를 찾았습니다. 모든 제한 조건 및 최적화 조건이 만족되었습니다."가 나온 것을 확인한다.
- 8)[확인] button을 click하면 완료된다.
- 9)결과값 cell B3에  $x = 0$ , cell C3에  $y = 1$ 을 확인한다.
- 10)이렇게 해서 앞에서 기술한  $x = 0$  일때,  $y$ 의 최소값은 1이 구해진다.

## Task 8. 선형 단순회귀분석 구하기

### 1. 목적

- Data 분석을 위해 수학 model을 작성한다.
- 이 model은 data를 담기 위한 변수와 구조를 결정하기 위한 '매개변수'로 설정이 되며, 이 매개변수를 결정하는 것이 '최적화'라고 부르는 수학적 기법이다.
- 이 최적화문제를 이해하기 위해 가장 알기 쉬운 예제가 '회귀분석'이다.
- 회귀분석 : 복수의 변수로 구성되는 자료에 대해서, 특정한 한 가지 변수에 주목하여 나머지 변수로 설명하는 기법
- 여러가지 회귀분석 방법이 있지만, 여기서는 가장 간단한 '선형 단순회귀분석'이라고 부르는 분석 방법을 살펴보자.
- '선형 단순회귀분석'은 2개의 변수로 구성되는 자료를 대상으로 한다.
- '선형 단순회귀분석'은 산점도 상의 점들을 대표하는 직선과 그 직선에서 두 변수 사이의 관계를 살펴보는 분석 기술이다('선형단순회귀분석.png'참조).
- 이 점들을 대표하는 직선을 '회귀직선'이라고 부른다.
- 회귀직선은 다음과 같이 1차 방정식으로 표현된다.
 
$$y = px + q(p, q \text{는 정수})$$
- 이것을 '회귀방정식'이라고 부른다.
- 여기서  $x$ 를 '설명변수',  $y$ 를 '목적변수'라고 부른다.
- 또한 정수  $p$ 와  $q$ 는 회귀분석 model을 결정하는 매개변수로, 주어진 data에 적합시키기 위해 적용된다.
- $p$ 를 '회귀계수',  $q$ 를 '절편'이라고 부른다.

### 2. 준비물

- 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task8] tab

### 3. 주어진 문제

- 1)다음 주어진 자료는 고등학교 3학년 여학생 7명의 신장과 체중 자료이다.
- 2)이 자료로부터 체중  $y$ 를 목적변수, 신장  $x$ 를 설명변수로 하는 회귀방정식  $y = px + q(p, q \text{는 정수})$ 를 구해본다.

197 4. 해결

- 198 1)가상의 매개변수  $p$ ,  $q$ 의 값을 입력(각각 1.00)하고, 이 값으로부터 회귀방정식( $y = px + q$ )을 이용하여 체중  $y$ 의 예측값( $=\$C\$3*\$C\$4+\$C\$4$ )을 계산한다.
- 199 2)제곱오차( $=(D7-E7)^2$ )를 산출한다.
- 200 3)제곱오차의 총합 QT를 SUM 함수로 산출한다.
- 201 4)[해 찾기] 메뉴를 선택하면 나오는 [해 찾기 매개 변수] 상자에서 아래와 같이 설정한다.
- 202 -목표 설정 :  $\$F\$14$
- 203 -대상 : 최소
- 204 -변수 셀 변경 :  $\$C\$3:\$C\$4$ (매개변수  $p$ ,  $q$ 의 주소를 설정)
- 205 -제한되지 않는 변수를 음이 아닌 수로 설정은 uncheck
- 206 5)[해 찾기] button을 click해서 실행하면, [해 찾기 결과] 상자가 나타나고, "모든 제한 조건을 만족시키는 해에 수렴했습니다"라는 메시지를 확인한 후, [확인] button을 click 한다.
- 207 6)그러면  $p$ 의 값 0.41,  $q$ 의 값 -11.97, 계 QT의 값 55.7의 결과가 생성된다.
- 208 7)이렇게 하면, 회귀계수와 절편  $p$ ,  $q$ 의 값이 구해진다.
- 209 8)또한 회귀방정식은 다음과 같이 나타낼 수 있다.
- 210  $y = 0.41x - 11.97$
- 211 9)위의 회귀방정식에 따라, 신장 170cm인 여학생의 체중을 예측해 보면 아래와 같다.
- 212 예측 체중  $y = 0.41 \times 170 - 11.97 = 57.73\text{kg}$
- 213
- 214

215 Task 9. 선형 단순회귀분석 구하기

216 1. 목적

- 217 -위의 Task8을 이용하여 선형 단순분석을 해보자.
- 218

219 2. 준비물

- 220 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task9] tab
- 221

222 3. 주어진 문제

- 223 1)다음 주어진 자료는 4개의  $x$ 의 값이다.
- 224 2)이 자료로부터  $x$ 의 5번째 값을 예측해보자.
- 225

226 4. 해결

- 227 1)예측값  $ax + b$ 의 값을 구한다.
- 228 2)실측값  $y$ 와 예측값  $ax + b$ 의 오차로, 제곱오차  $Q$ 를 구한다.
- 229 3)제곱오차  $Q$ 의 총합 QT를 산출한다.
- 230 4)[해 찾기]로 제곱오차의 총합 QT의 최소값을 구한다.
- 231 5)실행하면,  $a = 3.06$ ,  $b = 10.05$ 가 구해진다.
- 232 6)따라서 예측값  $y = 3.06x + 10.05$ 이다.
- 233 7)이 방정식에 따라  $x = 5$ 의 값을 대입하면 예측값 25.35가 산출된다.
- 234
- 235

236 Task 10. Microsoft Excel의 선형 단순회귀분석 구하기

237 1. 목적

- 238 -Microsoft Excel에서는 몇 개의 회귀분석 전용 도구가 준비되어 있다.
- 239
- 240

241 2. 준비물

- 242 1)Lab. 회귀분석과 최적화문제.xlsx 중의 [Task9] tab
- 243

244 3. 주어진 문제

245

246 4. 해결

- 247 1) Excel에서는 회귀계수를 구하기 위해 SLOPE함수가, 절편은 INTERCEPT라는 함수가 있다.
- 248 2) p의 값, 즉 C3는 =SLOPE(D7:D13,C7:C13)를 이용하고
- 249 3) q의 값, 즉 C4는 =INTERCEPT(D7:D13,C7:C13)를 이용한다.
- 250 4) 또한 회귀분석을 위한 전용도구도 있는데, [데이터] tab의 [데이터 분석] 메뉴를 click하면 [통계 데이터 분석] 상자가 열리는데, 여기서 [회귀분석]을 선택하면 된다.
- 251 5) 만일 [데이터 분석]이 설치되지 않았으면, 다음과 같이 실행한다.
- 252 -[파일] tab의 [옵션] 메뉴를 클릭한다.
- 253 -[Excel 옵션] 상자가 열리면 왼쪽 창 중에서 [추가 기능]을 선택한다.
- 254 -추가로 얻어진 box 중에서 아래에 있는 [관리] dropdown 메뉴 중 [Excel 추가 기능]을 선택하고, [이동] button을 click한다.
- 255 -[추가 기능] 상자가 열리면 [분석 도구]에 check하고 [확인] button을 click한다.
- 256 -설치 작업이 진행된다.
- 257 -바르게 설치되면 [데이터] 메뉴에 [데이터 분석] 항목을 확인할 수 있다.