5주차 읽기자료: ReLU 그래프와 MILP 유도

반응기 설계의 최적화 정식화

유량이 1 L/min으로 정해졌을 때 제품 규격을 만족하는 최소 부피를 찾습니다. 체류시간과 부피가 $V = Q\tau$로 연결되므로 τ 하나로 설계 선택을 표현합니다.

고정 조건은 330 K와 유입 A 농도 1.2 mol/L입니다. 모델 출력은 A·B·C의 농도이며 $B \geq 0.40$, $C \leq 0.25$ mol/L를 요구합니다.

먼저 공정 문제를 적고, 예측 계산을 제약식으로 바꾼 뒤, 계산된 해를 검증하는 순서로 풀어봅니다.

CSTR 설계 문제의 가능집합

$$\min_{\tau, y} V = Q\tau$$
$$2 \leq \tau \leq 10, y = f_\theta(\tau)$$
$$y_B \geq .40, y_C \leq .25, y \geq 0$$

τ의 범위는 모델을 사용할 영역이자 bounds 계산의 출발점입니다. 농도 부등식은 제품 요구사항을, y = fθ(τ)는 예측 모델의 계산을 나타냅니다.

목적함수의 V는 L 단위이고 제약의 y는 mol/L 단위입니다. τ와 V의 수치가 같은 이유는 Q = 1 L/min이기 때문입니다.

보간 자료와 모델 구성

τ [min]	A [mol/L]	B [mol/L]	C [mol/L]
2	0.857143	0.285714	0.057143
4	0.666667	0.380952	0.152381
6	0.545455	0.409091	0.245455
8	0.461538	0.410256	0.328205
10	0.400000	0.400000	0.400000

인접한 두 행 사이를 선형 보간합니다. 구간 안에서는 양 끝 농도의 convex combination이므로 비음수성과 농도 합 1.2가 유지됩니다.

표는 소수 여섯 자리로 표시했으며, 계산에는 원래 정밀도의 계수를 사용합니다.

선형 보간의 ReLU 표현 유도

각 knot의 농도 벡터를 c_i라고 두고 구간 기울기 s_i를 계산합니다. 첫 구간 기울기에서 출발해, knot를 지날 때 기울기 차이를 더합니다.

$$s_i = \frac{c_{i+1} - c_i}{\tau_{i+1} - \tau_i}$$

$$y(\tau) = c_0 + s_0(\tau - \tau_0) + \sum_{i=1}^{3} (s_i - s_{i-1}) \max(0, \tau - \tau_i)$$

구간 i 안에서 활성 hinge의 기울기를 합하면 $s_0 + (s_1 - s_0) + \cdots + (s_i - s_{i-1}) = s_i$가 됩니다. knot에서 연속이고 첫 값을 맞추므로 원하는 보간식입니다.

Scaling을 반영한 전체 가중치

$x = (\tau - 2)/8$, $h_j = \max(0, x - t_j)$, $t = [0, 0.25, 0.50, 0.75]$입니다. $y = b + W^{T}h$에서 W의 행은 뉴런, 열은 성분입니다.

j	W_jA	W_jB	W_jC
0	-0.761904762	0.380952381	0.380952381
1	0.277056277	-0.268398268	-0.008658009
2	0.149184149	-0.107892108	-0.041292041
3	0.089510490	-0.045687646	-0.043822844

$b = [0.857142857, 0.285714286, 0.057142857]$ mol/L입니다. W의 단위도 mol/L이며 h는 무차원입니다.

정확한 계수는 `frozen_network.json`에 저장됩니다. 표의 반올림값으로 다시 구성하면 작은 출력 차이가 생길 수 있습니다.

보간 기울기에서 ReLU 가중치로

B의 첫 세 절점 값은 τ=2,4,6에서 0.285714286,0.380952381,0.409090909입니다. x=(τ−2)/8이므로 절점의 x 간격은 2/8=0.25입니다.

계산	결과
첫 구간의 x 기울기	(0.380952381−0.285714286)/0.25=0.380952381
두 번째 구간의 x 기울기	(0.409090909−0.380952381)/0.25=0.112554113
첫 ReLU 가중치 W_0B	첫 기울기 0.380952381
절점 0.25의 가중치 W_1B	새 기울기−이전 기울기=−0.268398268

0.25<x<0.5에서는 첫 두 ReLU만 미분이 1이므로 전체 기울기는 0.380952381 − 0.268398268 = 0.112554113입니다. 새 항은 절점에서 0이므로 값의 연속성도 유지됩니다.

τ 기준 기울기를 사용했다면 x 기준 가중치로 옮길 때 8을 곱해야 합니다. 같은 농도 변화가 x에서는 8배 짧은 간격에 걸쳐 일어나기 때문입니다. A와 C 열도 같은 방법으로 구합니다.

τ=5 min에서 은닉값과 B 농도

x=(5−2)/8=0.375를 네 절점 [0,0.25,0.50,0.75]에서 뺍니다.

j	$a_j=x-t_j$	h_j	B에 더해지는 값 $W_jB\cdot h_j$
0	0.375	0.375	0.380952381×0.375=0.142857143
1	0.125	0.125	−0.268398268×0.125=−0.033549784
2	−0.125	0	0
3	−0.375	0	0

편향 0.285714286을 더하면 B≈0.395021645 mol/L입니다. 4분과 6분의 농도를 반씩 섞어도 (0.380952381+0.409090909)/2=0.395021645가 됩니다.

ReLU 가중합과 구간 보간이 같은 계산을 다른 표현으로 수행합니다. 기준 공정의 5분 B=0.4와는 차이가 있어, 이 예측은 B≥0.40 규격에 조금 못 미칩니다.

Big-M의 양방향 동치

$$h \geq 0, h \geq a$$
$$h \leq a - L(1 - z)$$
$$h \leq Uz, z \in \{0, 1\}$$

정식화의 가능점에서 z = 0이면 h = 0, a ≤ 0이고, z = 1이면 h = a ≥ 0입니다. 어느 분기에서도 h = max(0, a)입니다.

역으로 ReLU 그래프의 점을 택합니다. a ≤ 0이면 z = 0, a ≥ 0이면 z = 1로 두면 네 부등식이 모두 성립합니다. L ≤ a ≤ U의 유효성이 이 역방향에도 필요합니다.

하나의 ReLU 제약에 수치 대입

$a=2x-1$, $x\in[0,1]$이면 $L=-1, U=1$입니다. 네 식은 $h\ge 0$, $h\ge a$, $h\le a+1-z$, $h\le z$입니다.

입력과 가지	하한	상한	결정된 출력
$x=0.2$, $a=-0.6$, $z=0$	$h\ge 0$	$h\le 0$; $h\le 0.4$	$h=0$
$x=0.8$, $a=0.6$, $z=1$	$h\ge 0.6$	$h\le 0.6$; $h\le 1$	$h=0.6$
$x=0.8$에서 잘못된 $z=0$	$h\ge 0.6$	$h\le 0$	불가능

```
a = 2 * x - 1
h = max(0., a)
z = int(a >= 0.)
assert h >= 0 and h >= a
assert h <= a + 1 - z and h <= z
```

이진변수가 임의의 출력을 만드는 것이 아니라, 주어진 a와 양립하는 가지를 선택합니다. $a=0$에서는 두 z가 가능하지만 h는 모두 0입니다.

다층 ReLU 정식화의 동치성

입력 x를 고정합니다. 첫 affine layer는 a^1을 유일하게 정하고, 각 ReLU 제약은 h^1을 정확히 정합니다.

h^1이 정해지면 다음 affine layer의 a^2도 정해집니다. 같은 논리를 마지막 layer까지 반복하면 MILP의 출력은 forward 출력과 같습니다.

반대로 forward 계산의 모든 중간값을 넣고 각 뉴런의 부호로 z를 고르면 MILP 가능점이 됩니다. 전 layer에서 bounds가 유효해야 하며, 0인 뉴런에서 z가 여러 값이어도 출력은 같습니다.

Layer별 interval bounds의 전파

$$a = \sum_i w_i h_i + b$$
$$L = b + \sum_i \min\left(w_i l_i, w_i u_i\right)$$
$$U = b + \sum_i \max\left(w_i l_i, w_i u_i\right)$$

ReLU를 통과한 다음 구간은 [max(0, L), max(0, U)]입니다. 이 구간을 다음 affine layer의 입력 bounds로 전달합니다.

예를 들어 $h_1 = x$, $h_2 = x$이면 $h_1 - h_2$는 항상 0입니다. 각 h가 [0, 1]이라는 정보만 사용하면 [−1, 1]을 얻습니다. 이 범위도 유효하지만 상관관계를 놓쳤습니다.

목적함수와 제약 행렬의 변수 순서

변수 순서는 [x, h_0, h_1, h_2, h_3, y_A, y_B, y_C, z_1, z_2, z_3]입니다. 각 행은 row_lower ≤ Av ≤ row_upper 형태로 저장합니다.

V = 2 + 8x에서 상수 2는 최적해에 영향을 주지 않으므로 solver에는 8x만 전달합니다. 보고하는 V와 dual bound에는 2를 다시 더합니다.

정수변수의 bounds를 [0, 1]로 지정해야 이진변수가 됩니다. pre-activation을 별도 변수로 만들면 음수도 허용하도록 하한을 직접 지정합니다.

제약식 한 줄을 solver 행으로 저장하기

전체 변수 순서는 $[x, h_0, h_1, h_2, h_3, y_A, y_B, y_C, z_1, z_2, z_3]$입니다. $h_1 - x + 0.25 z_1 \le 0$의 계수는 해당 세 열에만 들어갑니다.

```
import numpy as np
row = [-1, 0, 1, 0, 0, 0, 0, 0, 0.25, 0, 0]
row_lower = -np.inf
row_upper = 0.0
```

대입	행 내적	판정
$x=0.5, h_1=0.25, z_1=1$	$-0.5+0.25+0.25=0$	상한 0 만족
$x=0.5, h_1=0.30, z_1=1$	$-0.5+0.30+0.25=0.05$	상한 위반

출력 등식 $y_B - \sum W_j B\ h_j = b_B$는 해당 행의 하한과 상한을 모두 b_B로 둡니다. z_1 열은 정수 표시와 $[0,1]$ bounds를 함께 설정합니다. 계수 배열만 만들고 이 두 설정을 빼면 이진변수가 되지 않습니다.

SciPy LinearConstraint는 각 행을 `lb <= A @ v <= ub`로 해석합니다. 따라서 변수 순서와 계수의 부호를 같은 행에서 대조합니다.

최적 체류시간 손계산

2-4 min 구간에서 B는 0.40에 못 미칩니다. 4-6 min 구간에서는 B가 증가하므로 최초로 0.40을 만족하는 점을 구합니다.

$$\tau = 4 + 2\frac{.40 - .380952380952}{.409090909091 - .380952380952}$$

$$\tau \approx 5.353846\text{min}$$

같은 보간 비율로 [A, B, C] = [0.584615, 0.400000, 0.215385] mol/L를 얻습니다. C ≤ 0.25와 비음수성을 만족하므로 이 점이 최소 체류시간입니다.

제품 하한에 도달하는 보간 비율

4~6분 구간에서 B는 0.380952381에서 0.409090909로 증가합니다. 필요한 증가량을 구간 전체 증가량으로 나눕니다.

계산	값
필요한 B 증가량	$0.40-0.380952381=0.019047619$
구간 전체 증가량	$0.409090909-0.380952381=0.028138528$
구간 비율 α	$0.019047619/0.028138528\approx0.676923077$
체류시간	$4+2\alpha\approx5.353846154$ min
같은 α의 C	$0.152380952+\alpha(0.245454545-0.152380952)\approx0.215384615$

이전 구간의 B는 모두 0.40보다 작습니다. 이 첫 교점에서 $C\le0.25$와 비음수 조건도 만족하므로 최소 부피를 얻습니다. B만 따로 보간한 뒤 다른 τ의 C와 비교하면 같은 설계의 규격 검사가 되지 않습니다.

스칼라 ReLU의 연속 완화

L < 0 < U에서 z ∈ [0, 1]로 풀고 z를 제거하면 다음 삼각형을 얻습니다.

$$h \geq 0, h \geq a$$
$$h \leq \frac{U}{U-L}(a - L)$$

두 상한에서 h/U ≤ z ≤ (h − a + L)/L을 얻습니다. L < 0의 부호 반전에 주의해 결합하면 마지막 부등식이 나옵니다.

이는 스칼라 (a, h) 그래프의 convex hull입니다. a가 여러 입력의 affine 함수일 때 원래 입력 공간의 가장 강한 완화라는 뜻은 아닙니다.

Affine 등식 투영의 MILP 표현

$$y = Py_{\text{raw}} + Ru + q$$

4주차의 투영이 위 affine 형태이고 P, R, q가 고정이면 출력 등식으로 그대로 추가할 수 있습니다. 추가 이진변수는 필요하지 않습니다.

제약 행렬이 조작변수에 따라 달라지면 역행렬이나 변수 곱이 생길 수 있으므로 같은 결론을 바로 적용할 수 없습니다.

보존식을 별도 제약으로 추가할 때는 기존 출력과 일치하는지도 확인합니다. 근사 예측에 새 등식을 추가하면 가능집합이 줄거나 사라질 수 있습니다.

연습 1: 부호와 분기 확인

$a = 2x - 1$, $x \in [0, 1]$에서 L과 U를 구하고 Big-M 식을 적으세요. $x = 0.2$와 0.8에서 (a, h, z)를 계산합니다.

확인: $L = -1$, $U = 1$이고 $h \geq 0$, $h \geq 2x - 1$, $h \leq 2x - z$, $h \leq z$입니다.

$x = 0.2$에서는 $(-0.6, 0, 0)$, $x = 0.8$에서는 $(0.6, 0.6, 1)$입니다. $x = 0.5$에서는 $h = 0$이고 z는 0 또는 1입니다.

연습 2: 제품 규격과 가능집합

$C \le 0.25$를 유지하면서 $B \ge 0.41$로 바꾸면 가능해가 남을까요? 표의 knot 값과 구간별 보간으로 판단하세요.

확인: 6 min에서 $C = 0.245455$이고 8 min에서 0.328205입니다. C 상한은 체류시간을 약 6.109859 min 이하로 제한합니다.

그 점의 B는 약 0.409155 mol/L이며 이전 구간의 B도 더 작습니다. 따라서 변경한 surrogate 문제는 불가능합니다. Solver의 infeasible 종료와 구현 오류는 이 계산으로 구별할 수 있습니다.

연습 3: 정식화 검증용 입력

난수 입력만 확인하면 꺾임점이나 영역 끝점의 오류를 놓칠 수 있습니다. $\tau = 2, 4, 6, 8, 10$과 내부 knot 양쪽의 점을 포함하세요.

같은 입력에서 h와 y의 최대 절대 차이를 계산합니다. $a = 0$에서 이진변수의 일치를 요구하지 말고 출력의 일치를 확인합니다.

확인: 제공 실습은 41개 입력의 최대 차이가 약 1.25×10^{-8}이었습니다. Solver 허용오차 때문에 knot에 매우 가까운 점에서 작은 차이가 남을 수 있습니다. 유효한 정식화의 수학적 동치와 유한 정밀도의 계산 결과를 구분합니다.

실습 파일과 출처

`week05_lab.py`는 계수 구성, MILP 행렬 조립, 고정 입력 검사와 구간별 독립 계산을 포함합니다. `frozen_network.json`과 `fit_knots.csv`는 모델을, `results.json`과 `milp_matrix.csv`는 계산을 확인할 때 사용합니다.

Boyd & Vandenberghe, Convex Optimization, 4.1절: 변수·목적함수·가능집합.

Anderson et al., Strong mixed-integer programming formulations for trained neural networks, 2.1절: ReLU Big-M 표현.

SciPy milp 공식 문서: bounds, integrality, LinearConstraint와 반환 상태.

강의노트 · 실습 묶음