7주차: 최적화 활용을 위한 모델 구조 설계 | 해설과 연습

가열 부하 선택의 수학적 범위

가열 부하 u를 [0, 1]에서 선택할 때 사용한 비용은 무차원 계산 예제입니다. 실제 공정의 열역학 모델에서 convexity를 유도한 비용식은 아닙니다.

공급 조건을 고정한 비용의 형상이 convex라는 근거가 있을 때 구조 제약을 검토합니다. 농도·수율·회수율에는 이 가정을 자동으로 적용하지 않습니다.

네트워크의 convexity, 정식화의 동치성, 기준 함수에서의 의사결정 품질을 각각 확인합니다.

Convexity 증명의 귀납 단계

h의 각 성분이 convex이고 $W \geq 0$이라고 가정합니다. $0 \leq \lambda \leq 1$에서 성분별 convexity 부등식에 W를 곱해도 부등식 방향이 유지됩니다.

$$Wh(\lambda u + (1-\lambda)v) \leq \lambda Wh(u) + (1-\lambda)Wh(v)$$

Affine 항 Au + b에는 등식이 성립합니다. 두 항을 더한 뒤 convex·비감소 ReLU를 합성하면 다음 층도 convex입니다. 같은 논증을 출력의 비음수 가중합에 적용합니다.

확장 epigraph의 두 포함관계

실제 forward 출력을 h^*라 둡니다. 확장 제약의 h는 첫 층에서 $h \geq h^*$입니다. 이전 층의 이 관계와 $W \geq 0$을 다음 층 제약에 대입하면 같은 부등식이 이어집니다.

따라서 모든 확장 feasible point에서 $t \geq f(u)$입니다. 반대로 $t \geq f(u)$인 임의의 (u, t)에 $h = h^*$를 붙이면 확장 제약을 만족합니다.

즉, 보조변수를 제거한 (u, t) 집합은 정확히 epigraph입니다. 이 증명은 모든 최적 보조변수가 반드시 tight하다고 요구하지 않습니다.

두 층의 보조변수를 forward로 복원하기

$h_1=\mathrm{ReLU}(u-0.4)$, $h_2=\mathrm{ReLU}(2h_1-0.1)$, $f=0.2-u+3h_2$인 ICNN을 봅니다. 은닉 연결 2와 출력 연결 3이 양수입니다. 입력 $u=0.6$을 고정합니다.

상태	h_1	h_2	출력식의 값
실제 forward	0.2	0.3	0.5
느슨한 epigraph 가능점	0.4	0.7	1.7
첫 층만 실제값으로 내림	0.2	0.7	1.7
둘째 층도 실제값으로 내림	0.2	0.3	0.5

h_1을 낮추면 $2h_1-0.1$도 낮아져 기존 h_2의 하한 제약을 깨뜨리지 않습니다. 이어 h_2를 낮추면 양의 출력 계수 때문에 목적값도 내려갑니다. 이 순서가 층별 복원 논증입니다.

출력 계수가 -3이면 h_2를 높일수록 목적값이 낮아져 이 논리가 성립하지 않습니다. 외부에서 $h_2\ge 0.7$을 요구해도 실제값 0.3으로 복원할 수 없으므로, 부호와 외부 제약을 함께 확인합니다.

목적함수 단조성과 보조변수 복원

목적함수가 t에 대해 비감소이고, 나머지 제약도 t와 h를 forward 값까지 줄이는 것을 허용하면 최적값을 보존하며 복원할 수 있습니다.

min t에서는 이 조건이 직접 성립합니다. $t \le b$인 상한도 줄이는 방향을 허용합니다. 반면 $h \ge b$, 출력 보상, 출력 등식과 결합된 다른 상태 제약은 따로 검사해야 합니다.

출력에 연결되지 않은 뉴런의 w가 0이라면 h에 slack이 남아도 목적값은 바뀌지 않습니다. 해석에는 u를 다시 forward 계산한 출력을 사용합니다.

두 hinge 모델의 구간별 계산

u 구간	f(u)	기울기
[0, 0.4]	$0.4 - 0.6u$	-0.6
[0.4, 0.7]	$0.4u$	0.4
[0.7, 1]	$0.9u - 0.35$	0.9

연속성을 확인하면 $f(0.4) = 0.16$, $f(0.7) = 0.28$입니다. 기울기가 음수에서 양수로 바뀌는 0.4가 유일한 최소점입니다.

$f \le 0.25$를 각 구간에서 풀면 $u \ge 0.25$와 $u \le 0.625$가 남습니다. 세 번째 구간에는 해가 없습니다.

두 절점 사이의 비용과 가능 구간

각 구간에서 켜진 ReLU만 남겨 비용식을 정리합니다.

u 구간	켜진 항	비용식	f(u)≤0.25의 조건
0~0.4	없음	0.4−0.6u	u≥0.25
0.4~0.7	u−0.4	0.4u	u≤0.625
0.7~1	두 항 모두	0.9u−0.35	u≤2/3, 해당 구간에 없음

가능 구간은 [0.25,0.625]입니다. 최소점 u=0.4에서 두 ReLU가 0이므로 f=0.4−0.6×0.4=0.16입니다.

예를 들어 u=0.6이면 h_1=0.2,h_2=0이고 f=0.4−0.36+0.2=0.24입니다. 중간 구간의 0.4u=0.24와 일치하므로 구간식과 원래 신경망을 서로 확인할 수 있습니다.

SciPy에 전달하는 LP 행렬

변수 순서는 [u, h_1, h_2]이고 목적함수 상수 0.4는 풀이 뒤 더합니다.

```
result = linprog(
    [-0.6, 1.0, 0.5],
    A_ub=[[1, -1, 0], [1, 0, -1]],
    b_ub=[0.4, 0.7],
    bounds=[(0, 1), (0, None), (0, None)],
    method='highs')
value = result.fun + 0.4
```

행렬의 첫 행은 $u - h_1 \le 0.4$입니다. 반환된 status를 확인한 뒤 result.x를 사용합니다.

LP 행렬의 열과 부등식

변수 순서를 $[u,h_1,h_2]$로 고정합니다. linprog의 A_ub는 각 행과 변수 벡터를 내적한 값이 b_ub 이하가 되도록 해석합니다.

코드의 값	해당 수식	역할
목적 계수 $[-0.6,1,0.5]$	$-0.6u+h_1+0.5h_2$	상수 0.4를 뺀 비용
행 $[1,-1,0]$, 우변 0.4	$u-h_1\le 0.4$	$h_1\ge u-0.4$
행 $[1,0,-1]$, 우변 0.7	$u-h_2\le 0.7$	$h_2\ge u-0.7$
h의 bounds (0,None)	$h_1,h_2\ge 0$	ReLU의 다른 하한

해 $[0.4,0,0]$을 대입하면 행 값은 $[0.4,0.4]$로 우변 이하입니다. solver 목적값 -0.24에 상수 0.4를 더해 0.16을 보고합니다. 원래 forward에서 계산한 비용도 0.16이어야 합니다.

양의 h 계수 때문에 이 목적식은 h를 가능한 작게 만듭니다. 목적 계수의 부호를 바꾸면 이 복원 논리를 다시 검사해야 합니다.

부등식 방향 반례의 완전한 해

실제 문제는 $\min u,\ 0 \le u \le 1,\ \mathrm{ReLU}(u) \ge 0.8$입니다. 이 입력 구간에서 $\mathrm{ReLU}(u) = u$이므로 최적값은 0.8입니다.

완화에서는 $h \ge u,\ h \ge 0,\ h \ge 0.8$을 사용합니다. 모든 $u \in [0, 1]$에 $h = \max(u, 0.8)$를 붙일 수 있어 feasible 입력이 전체 구간으로 늘어납니다.

최적값은 0입니다. Solver는 변경된 문제를 정확히 풀었지만 원래 문제의 답은 얻지 못했습니다.

고정 knot ICNN의 학습 행렬

$$\Phi_i = \left[1, u_i, \mathrm{ReLU}\left(u_i - \kappa_1\right), \ \ldots, \mathrm{ReLU}\left(u_i - \kappa_9\right)\right]$$

$$\beta = \left[b, a, w_1, \ \ldots, w_9\right]^T$$

$$\min_\beta \left\| \Phi\beta - q \right\|_2^{\,2}, w \geq 0$$

b와 a는 자유롭고 w만 비음수입니다. Φ를 고정하면 목적함수의 Hessian은 $2\Phi^T\Phi$로 positive semidefinite이므로 convex quadratic fit입니다.

Knot 위치나 깊은 층의 가중치도 함께 학습하면 Φ가 매개변수에 의존해 이 논증을 그대로 적용할 수 없습니다.

예를 들어 u=0.25의 특성 행은 [1,0.25,0.15,0.05,0,0,0,0,0,0,0]입니다. 절점 0.1과 0.2만 지났으므로 두 hinge만 양수입니다. b,a 두 계수는 음수도 허용하고 나머지 9개에는 하한 0을 주어 최소제곱을 풉니다.

기준 최소점과 결정 손실

$$q(u) = (u - .6)^2 + .1u + .2$$
$$q'(u) = 2(u - .6) + .1 = 0 \Rightarrow u^* = .55$$
$$q(.5) - q(.55) = .2600 - .2575 = .0025$$

최소점 부근에서 비용 곡선이 완만하면 작은 예측 오차가 입력 위치를 바꿀 수 있습니다. 입력 오차 0.05와 비용 손실 0.0025는 서로 다른 평가량입니다.

학습한 기울기 중 [0.5, 0.6] 구간은 약 0.0000233으로 거의 평평합니다. solver 허용오차를 함께 확인합니다.

예측 오차와 선택한 입력의 손실

기준 비용 $q(u)=(u-0.6)^2+0.1u+0.2$를 완전제곱으로 정리하면 $q(u)=(u-0.55)^2+0.2575$입니다. 최소점에서 벗어난 거리의 제곱이 비용 증가량입니다.

입력	기준 비용 계산	비용
기준 최소 $u=0.55$	$0^2+0.2575$	0.2575
모델 선택 $u=0.50$	$(-0.05)^2+0.2575$	0.2600
차이	$0.2600-0.2575$	0.0025

모델이 선택한 점의 예측 비용 0.258437과 기준 비용 0.260000의 차이 0.001563은 그 점의 예측 오차입니다. 기준 최소값과의 차이 0.0025는 선택 때문에 생긴 비용 손실입니다.

같은 숫자를 비교하는 기준이 다릅니다. 정확도 표와 의사결정 평가 표에서 각각 무엇을 빼는지 명시합니다.

PICNN 계수 고정과 context gate

Context 경로 r(c)가 gate s(c) ≥ 0을 계산한다고 합시다. 고정된 c에서 W diag(s(c))가 비음수이면 이전 convex 활성화의 비음수 가중합을 유지합니다.

u에 곱하는 gate는 고정된 수이므로 부호가 자유로운 affine 입력 연결에 포함할 수 있습니다.

실습에서는 $|u - c| - 0.1c^2$로 고정 context에서의 LP와 공동 convexity 반례를 확인합니다. 원 논문의 전체 PICNN 구조를 학습하는 실험은 포함하지 않습니다.

고정 c=0.2에서는 $|u - 0.2|$를 보조변수 t의 두 하한 $t \geq u - 0.2$, $t \geq 0.2 - u$로 표현하고 $t - 0.004$를 최소화합니다. $u = 0.2, t = 0$에서 최솟값 -0.004입니다. c가 변수가 되면 목적식의 $-0.1c^2$가 상수가 아니므로 이 LP를 그대로 사용할 수 없습니다.

Scaling과 물리 제약의 관계

고정된 affine 입력 변환 u = Sx + r에 convex f를 합성하면 x에 대해서도 convex입니다. 하지만 bounds, 단위와 고정 context는 정식화와 forward 코드에서 같아야 합니다.

양의 출력 scale은 convexity를 보존합니다. 음의 scale은 convex 함수를 concave로 바꾸므로 목적함수·부등식 방향을 함께 바꿔 해석해야 합니다.

농도 합을 맞추는 projection은 성분을 음의 계수로 섞을 수 있습니다. Week 4의 등식 보장만으로 각 출력의 convexity나 이 LP 표현을 보장하지는 않습니다.

연습 문제

1. $f(u) = \mathrm{ReLU}(-2u + 1)$의 입력 convexity를 판정하고 근거를 적습니다.
2. 두 hinge 모델에서 $f \le 0.20$의 feasible 구간을 구합니다.
3. $h \ge \mathrm{ReLU}(u)$를 나타내는 선형 부등식에 $h \le 1$을 추가하면 $h = \mathrm{ReLU}(u)$가 되는지 반례를 찾습니다.
4. $f(u;c) = |u - c| - 0.1c^2$의 공동 convexity를 중점으로 검사합니다.
5. 비용 모델의 독립 RMSE가 줄어들 때 최적 입력 오차도 반드시 줄어드는지 설명합니다.

연습 문제의 짧은 답

1. Convex입니다. Affine 함수 $-2u + 1$과 0의 최댓값입니다.
2. $[1/3, 1/2]$입니다. 왼쪽에서는 $0.4 - 0.6u \le 0.2$, 가운데에서는 $0.4u \le 0.2$를 풉니다.
3. $u = 0, h = 0.8$이 반례입니다. 상한만으로 그래프 등식을 얻지 못합니다.
4. $(u, c) = (0, 0), (1, 1)$의 평균 함수값은 -0.05, 중점 값은 -0.025입니다.
5. 보장되지 않습니다. 최적점 주변의 국소 형상·기울기와 제약 경계의 오차가 결정에 영향을 줍니다.

재현 기록과 검토 기준

실습은 NumPy 2.5.4, SciPy 1.18.1에서 실행했습니다. 계산시간은 실행 환경에 따라 달라집니다.

제출물에는 학습 계수, 데이터 구분, 부호 검사, LP status와 forward 오차, 기준 비용 손실을 포함합니다. 최소점이 여러 개이면 목적값 동치성과 입력의 비유일성을 구분합니다.

Amos, Xu and Kolter, Input Convex Neural Networks, ICML 2017, pp. 146-155

Boyd and Vandenberghe, Convex Optimization, §§3.1-3.2 and 4.1-4.3