‘cat’은 어느 토큰을 읽을 수 있을까?

각 행은 토큰 하나입니다. ‘cat’ 위치의 질의로 계산하세요. 키의 차원은 2, 점수 온도는 1입니다. 계산기만 있으면 됩니다. `q(cat) = [1, 1]`

> $s_i = (q \cdot k_i) / \sqrt{2}$
> $w_i = \exp(s_i) / \Sigma_j \exp(s_j)$

1. 두 가지 가중치 행을 채워 보세요

먼저 네 키를 모두 사용합니다. 이어서 인과 마스크를 켜면 ‘cat’은 자신과 앞선 위치만 읽습니다. 분모에는 보이는 키만 포함하고 가려진 키의 가중치는 0으로 적으세요.

토큰	K	V	q · k	전체 가중치	인과 가중치
The	[1, 0]	[1, 0]			
cat	[1, 1]	[0, 2]			
sat	[0, 1]	[2, 1]			
here	[-1, 0]	[-1, 1]			

2. 값 벡터를 섞어 보세요

각 V에 그 위치의 가중치를 곱해 네 벡터를 더하세요. 마스크를 껐을 때와 켰을 때의 출력 벡터를 각각 적으세요.

3. 미래 위치의 값 하나를 바꿔 보세요

V(here)에 [0, 8]을 더합니다. 앞에서 구한 가중치만 이용해 두 경우의 출력 변화량을 구하세요.

4. 마스크 처리 오류를 찾아보세요

동료가 네 점수에 소프트맥스를 적용한 뒤 마지막 두 가중치를 0으로 바꿨습니다. 어느 숫자를 확인하면 오류를 찾을 수 있을까요?

© 2026 Seongeun So · github.com/sungeuns/foundation-model-engineering · MIT 라이선스

풀이

내적을 각각 $\sqrt{2}$로 나눕니다. 인과 마스크를 켜면 처음 두 토큰만 보이고, 두 점수는 약 0.7071과 1.4142입니다. 보이는 토큰만 분모에 넣어 각 가중치를 구합니다.

> `w(The) = exp(1/√2) / (exp(1/√2) + exp(2/√2)) ≈ 0.3302`

토큰	K	V	q · k	전체 가중치	인과 가중치
The	[1, 0]	[1, 0]	1	0.2341	0.3302
cat	[1, 1]	[0, 2]	2	0.4748	0.6698
sat	[0, 1]	[2, 1]	1	0.2341	0.0000
here	[-1, 0]	[-1, 1]	-1	0.0569	0.0000

표의 가중치는 반올림했습니다. 최종 출력은 반올림 전 값으로 계산합니다.

2. 값 벡터를 섞어 보세요

가중치 행을 w라 하면 첫 좌표는 w(The) + 2w(sat) − w(here), 둘째 좌표는 2w(cat) + w(sat) + w(here)입니다.

> 전체 키 → [0.6455, 1.2407]
> 인과 마스크 → [0.3302, 1.3395]

3. 미래 위치의 값 하나를 바꿔 보세요

값 벡터만 바꿨으므로 질의와 키는 그대로이고 가중치도 변하지 않습니다. 둘째 출력 좌표만 변할 수 있습니다. 전체 키를 쓰면 그 좌표는 8 × w(here) ≈ 0.4554만큼 커집니다. 인과 마스크를 켜면 그 가중치가 0이므로 두 출력 좌표가 모두 그대로입니다.

4. 마스크 처리 오류를 찾아보세요

뒤늦게 가리면 살아남은 두 가중치의 합은 0.7090입니다. 정규화한 행의 합은 1이어야 합니다. 소프트맥스 전에 점수를 가리거나, 남은 가중치의 합이 양수일 때 그 합으로 각각 나눕니다.

브라우저에서 다른 질의도 실험하기 · 같은 계산을 하는 PyTorch 예제

© 2026 Seongeun So · github.com/sungeuns/foundation-model-engineering · MIT 라이선스