

TensorFlow Playground XOR 실험 분석: 성공과 실패 사례 비교

XOR(배타적 논리합) 문제는 머신러닝에서 비선형 문제를 해결하는 능력을 테스트하는 고전적인 예시입니다. 입력값이 서로 다를 때만 참(1)을 반환하므로, 단일 직선으로는 데이터를 완벽하게 분류할 수 없습니다.

이 문서는 텐서플로우 플레이그라운드를 활용하여 XOR 데이터를 성공적으로 학습하는 경우와 실패하는 경우의 하이퍼파라미터 설정을 비교하고, 그 원인을 심층적으로 분석합니다.

실험 1: 학습 실패 사례 (낮은 성능)

이 시나리오는 잘못된 설정으로 인해 모델이 XOR 패턴을 전혀 학습하지 못하는 경우를 보여줍니다.

◆ 설정값

- **Problem type:** Classification
- **Features:** x_1, x_2 (기본 특징만 사용)
- **Hidden Layers:** 1개 층, 2개 뉴런
- **Activation:** Linear (비선형성 없음)
- **Learning Rate:** 0.3 (과도하게 높음)
- **Regularization:** L2, $\lambda=0.1$ (상대적으로 강한 규제)

◆ 결과 및 분석

결정 경계가 단순한 직선 형태로 나타나며, 네 개의 데이터 군집 중 두 개만 겨우 맞추는 수준에 그칩니다. 결과적으로 **Train/Test Loss는 0.4 이상으로 높게 유지되고, 정확도는 50% 수준에 머무릅니다.**

이러한 실패의 핵심 원인은 다음과 같습니다.

1. **Linear 활성화 함수:** XOR 문제는 본질적으로 비선형 경계가 필요합니다. Linear 함수는 입력값을 그대로 출력하므로, 여러 층을 쌓아도 결국 하나의 선형 함수처럼 동작합니다. 따라서 비선형 패턴을 절대 학습할 수 없습니다.
2. **부족한 모델 복잡성:** 은닉층이 1개이고 뉴런 수가 2개에 불과하여, 복잡한 패턴을 표현하기에는 모델의 용량 (capacity)이 절대적으로 부족합니다.
3. **높은 학습률 (Learning Rate):** 학습률이 너무 높으면 가중치(weight)가 최적의 값으로 수렴하지 못하고 발산하거나, 최적점 주변에서 크게 진동하여 안정적인 학습이 불가능합니다.
4. **과도한 규제 (Regularization):** L2 규제는 모델의 가중치가 너무 커지지 않도록 제한하여 과적합을 방지하는 역할을 합니다. 하지만 이 경우처럼 모델이 단순한 상황에서 규제를 강하게 적용하면, 오히려 패턴 학습을 더욱 방해하여 모델이 지나치게 단순화됩니다.

실험 2: 학습 성공 사례 (높은 성능)

올바른 하이퍼파라미터 설정을 통해 XOR 문제를 완벽하게 해결하는 경우입니다.

◆ 설정값

- **Problem type:** Classification
- **Features:** x_1 , x_2 , x_1x_2 , x_1^2 , x_2^2 (파생 특징 추가)
- **Hidden Layers:** 2개 층, 각 층에 4~8개 뉴런
- **Activation:** Tanh 또는 ReLU (비선형 함수)
- **Learning Rate:** 0.03 (안정적인 값)
- **Regularization:** 없음 ($\lambda=0$)

◆ 결과 및 분석

결정 경계가 데이터의 네 사분면을 정확하게 감싸는 비선형 형태로 나타납니다. 그 결과 **Train/Test Loss는 0.01 미만으로 안정적으로 감소하며, 정확도는 95% 이상**을 달성합니다.

성공의 핵심 요인은 실패 사례와 정확히 대조됩니다.

5. **Tanh / ReLU 활성화 함수**: Tanh나 ReLU 같은 비선형 활성화 함수는 모델이 복잡한 곡선 형태의 결정 경계를 만들 수 있게 해줍니다. 이것이 XOR 문제 해결의 가장 핵심적인 요소입니다.
6. **충분한 모델 복잡성**: 2개 이상의 은닉층과 각 층의 충분한 뉴런은 XOR와 같이 복잡한 함수를 근사할 수 있는 충분한 용량을 제공합니다.
7. **적절한 학습률 (Learning Rate)**: 학습률이 안정적이므로, 손실(Loss)이 발산하지 않고 부드럽게 감소하며 최적의 가중치를 찾아갑니다.
8. **파생 특징(Feature)의 활용**: x_1x_2 와 같은 교차항 특징은 모델이 데이터의 상호작용을 더 쉽게 학습하도록 돕습니다. 이 특징은 XOR 문제에서 특히 강력한 힌트 역할을 하여 학습 속도와 성능을 크게 향상시킵니다.

III 설정별 성능 비교 요약

구분	나쁜 설정 (성능 낮음)	좋은 설정 (성능 높음)	이유 분석
활성화 함수	Linear	Tanh / ReLU	XOR는 비선형 문제이므로, Linear 함수로는 학습이 불가능하다.
은닉층/뉴런	1층, 뉴런 2개	2층, 각 4~8 뉴런	모델의 깊이와 너비가 충분해야 복잡한 결정 경계를 만들 수 있다.
학습률	0.3 (너무 큼)	0.03 (적절)	큰 값은 가중치 발산을, 적절한 값은 안정적인 수렴을 유도한다.
사용 특징	x_1, x_2	$x_1, x_2, x_1x_2, x_1^2, x_2^2$	x_1x_2 와 같은 파생 특징은 비선형 관계 학습에 결정적인 도움을 준다.
규제	L2 ($\lambda=0.1$)	없음 ($\lambda=0$)	단순한 모델에 강한 규제를 걸면 오히려 필요한 패턴 학습을 방해한다.

결론: XOR 문제 해결의 핵심

텐서플로우 플레이그라운드를 통한 XOR 실험은 딥러닝 모델의 성공적인 학습을 위해 다음 세 가지 요소가 얼마나 중요한지 명확하게 보여줍니다.

9. **비선형성 확보**: 문제에 맞는 **활성화 함수**를 선택하는 것은 가장 기본적이면서도 결정적인 요소입니다.
10. **적절한 모델 용량**: 풀고자 하는 문제의 복잡도에 맞춰 ****충분한 깊이(은닉층)와 너비(뉴런)****를 설계해야 합니다.
11. **안정적인 학습 환경**: **학습률, 규제, 특징 공학** 등 다양한 하이퍼파라미터를 신중하게 조정하여 최적의 학습 조건을 만들어야 합니다.