

TensorFlow Playground: Circle 데이터셋

TensorFlow Playground는 신경망 모델의 여러 설정(하이퍼파라미터)이 학습 결과에 어떤 영향을 미치는지 직관적으로 보여주는 훌륭한 도구입니다. 이 가이드에서는 비선형(non-linear) 데이터의 대표 주자인 **Circle 데이터셋**을 이용해, 모델의 성능이 극과 극으로 나뉘는 두 가지 사례를 비교 분석합니다.

☛ 실패 사례: 왜 모델은 원을 그리지 못할까?

먼저, 모델이 원형 패턴을 전혀 학습하지 못하고 성능이 나쁘게 나오는 경우를 살펴보겠습니다.

⚙️ 설정값

- **Features:** X_1, X_2 (기본 좌표만 사용)
- **Hidden Layers:** 1개 층, 2개 뉴런 (매우 단순한 구조)
- **Activation:** Linear (선형)
- **Learning Rate:** 0.3 (지나치게 높음)
- **Regularization:** L2, Rate = 0.1 (다소 강한 규제)

📊 결과 분석

- **결정 경계(Decision Boundary):** 데이터의 원형 분포를 무시하고 거의 직선에 가깝게 형성됩니다.
- **학습 성능:** Train loss와 Test loss 모두 높은 상태에 머무르며, 정확도가 매우 낮습니다. Loss 값이 안정적으로 줄어들지 않고 크게 진동하거나 발산할 수 있습니다.
- **결론:** 모델이 데이터의 핵심 패턴인 '원'을 전혀 인지하지 못했습니다.

실패 원인

1. **치명적인 활성화 함수 선택 (Activation: Linear):** Linear 함수는 여러 층을 쌓아도 결국 하나의 직선(또는 평면)으로 표현됩니다. 마치 자(ruler)만 가지고 곡선을 그리려는 것과 같아서, 원형 패턴 학습이 근본적으로 불가능합니다.
2. **과도하게 높은 학습률 (Learning Rate: 0.3):** 학습률이 너무 높으면 모델이 최적의 가중치를 찾아가는 과정에서 정답을 지나쳐 버리거나, 오히려 손실(loss)이 커지는 방향으로 튕겨 나갑니다. (발산)
3. **단순한 네트워크 구조 (1 Hidden Layer, 2 Neurons):** 설령 비선형 활성화 함수를 썼더라도, 은닉층과 뉴런의 수가 너무 적으면 복잡한 원형 패턴을 표현할 표현력이 부족합니다.
4. **불필요한 규제 (Regularization):** 모델이 너무 단순해서 과적합(overfitting)이 일어날 가능성이 없는데도 강한 규제를 적용했습니다. 이는 모델이 그나마 있던 작은 패턴조차 학습하지 못하게 방해하는 족쇄 역할을 합니다.

👉 성공 사례: 완벽한 원을 만드는 비결

이번에는 설정을 최적화하여 원형 경계를 완벽하게 학습하는 성공 사례를 보겠습니다.

⚙️ 설정값

- **Features:** X_1 , X_2 , X_1^2 , X_2^2 , $\sin(X_1)$, $\sin(X_2)$ (파생 특징 적극 활용)
- **Hidden Layers:** 2개 층, 각 8개, 6개 뉴런 (충분히 깊고 넓은 구조)
- **Activation:** Tanh (비선형)
- **Learning Rate:** 0.03 (안정적인 값)
- **Regularization:** 없음 (None)

결과 분석

- **결정 경계(Decision Boundary):** 데이터 분포를 정확히 따라가는 아름다운 원형 경계가 형성됩니다.
- **학습 성능:** Train loss와 Test loss가 0에 가깝게 안정적으로 수렴하며, 정확도는 98% 이상으로 매우 높습니다.
- **결론:** 모델이 데이터의 비선형 패턴을 완벽하게 이해하고 일반화했습니다.

성공 비결

5. **똑똑한 특징 공학 (Feature Engineering):** 가장 결정적인 성공 요인입니다. X_1^2 , X_2^2 같은 특징을 추가하는 것은 모델에게 "이 데이터는 원점으로부터의 거리($x^2+y^2=r^2$)와 관련이 있을 수 있어!" 라고 직접적인 힌트를 주는 것과 같습니다. 이 힌트 덕분에 모델은 복잡한 계산 없이도 원형 경계를 훨씬 쉽고 빠르게 찾아낼 수 있습니다.
6. **적절한 비선형 활성화 함수 (Activation: Tanh):** Tanh나 ReLU 같은 비선형 활성화 함수는 모델이 직선을 벗어나 부드러운 곡선 형태의 경계를 학습할 수 있게 해줍니다. 원형 패턴 학습에 필수적입니다.
7. **충분한 네트워크 용량 (2 Hidden Layers, 8/6 Neurons):** 여러 층과 충분한 뉴런은 모델이 입력된 특징들을 다양하게 조합하여 복잡한 패턴을 표현할 수 있는 충분한 '작업 공간'을 제공합니다.
8. **안정적인 학습률 (Learning Rate: 0.03):** 너무 크지도, 작지도 않은 적절한 학습률은 모델이 최적점을 향해 빠르고 안정적으로 나아가게 합니다.

★ 최종 정리: 성공과 실패의 핵심 차이

항목 (Item)	👎 나쁜 설정	👍 좋은 설정	핵심 이유
활성화 함수	Linear	Tanh / ReLU	비선형 문제에는 비선형 활성화 함수 가 필수!
네트워크 구조	얕고 좁게	깊고 넓게	복잡한 패턴을 표현할 충분한 공간 이 필요.
학습률	너무 높음	적절함	안정적인 학습을 위한 적절한 보폭 설정.
특징 (Features)	기본만 사용	파생 특징 추가	가장 강력한 힌트 를 제공하여 학습을 도움.
규제	강함	없음	단순한 모델에겐 규제가 오히려 학습 방해 요소 .

결론: 비선형(Non-linear) 문제를 풀기 위해서는 그에 맞는 비선형 도구(활성화 함수, 충분한 네트워크)와 좋은 재료(특징 공학)가 반드시 필요합니다.