MODÜL 6: Büyü Bozumu (Backpropagation)

Bu kısım, yapay zeka tarihinin en büyük buluşudur. Model bir tahmin yaptı ve hata oluştu. Şimdi şu soruyu soruyoruz: **"Bu hatanın ne kadarı W2'nin suçu, ne kadarı W1'in suçu?"**

Buna **"Hata Ataması" (Credit Assignment)** denir.

🧠 Mantık: Hatayı Geriye Doğru Akıtmak

Zincir kuralını hatırlıyor musun? Soğanı dıştan içe soyuyorduk. Şimdi iki katmanlı soğanımız var.

Yolculuğumuz şöyle (Tersten):

$$Loss \leftarrow Tahmin(A_2) \leftarrow Z_2 \leftarrow \mathbf{W_2} \leftarrow Gizli(A_1) \leftarrow Z_1 \leftarrow \mathbf{W_1}$$

Bunu kodlarken iki aşamaya böleceğiz:

1. **Aşama 1 (Çıktı Katmanı):** Hatayı W2’'ye yükle.
2. **Aşama 2 (Gizli Katman):** Kalan hatayı W2 üzerinden geriye (W1'e) taşı.

1. Hedef: Çıkıştaki Hata (dZ2) Neden `p - Y`?

Amacımız şu türevi bulmak: Loss fonksiyonunun, Sigmoid'in girdisine (Z2) göre değişimi.

$$\frac{\partial Loss}{\partial Z_2} = ?$$

Zincir kuralı der ki:

$$\frac{\partial Loss}{\partial Z_2} = \underbrace{\frac{\partial Loss}{\partial p}}_{\text{Adım A}} \cdot \underbrace{\frac{\partial p}{\partial Z_2}}_{\text{Adım B}}$$

ADIM A: Loss Fonksiyonunun Türevi

Fonksiyonumuz (Binary Cross Entropy):

$$Loss = -[y \cdot \ln(p) + (1-y) \cdot \ln(1-p)]$$

Bunun p'ye göre türevini alalım.

- $\ln(p)$'nin türevi $\frac{1}{p}$'dir.
- $\ln(1-p)$'nin türevi $\frac{-1}{1-p}$'dir (Zincir kuralından gelen -1'e dikkat).

Türev denklemi:

$$\frac{\partial Loss}{\partial p} = -\left[\frac{y}{p} - \frac{1-y}{1-p}\right]$$

Paydaları eşitleyelim (Matematiksel jimnastik):

$$= -\left[\frac{y(1-p) - p(1-y)}{p(1-p)}\right]$$

$$= \frac{-y + yp + p - py}{p(1-p)}$$

(Burada $+yp$ ve $-py$ birbirini götürür)

$$= \frac{p-y}{p(1-p)}$$

Elimizde bu var: $\frac{p-y}{p(1-p)}$ (Bu haliyle çok çirkin duruyor, bekle).

ADIM B: Sigmoid Fonksiyonunun Türevi

Fonksiyonumuz: $p = \sigma(Z_2) = \frac{1}{1+e^{-Z_2}}$

Sigmoid fonksiyonunun türevinin çok özel bir özelliği vardır. Uzun uzun türev alırsan (bölüm türevi kuralıyla) şu harika sonuca ulaşırsın:

$$\sigma'(z) = \sigma(z) \cdot (1 - \sigma(z))$$

Yani bizim dilimizde:

$$\frac{\partial p}{\partial Z_2} = p \cdot (1 - p)$$

ADIM C: BÜYÜK FİNAL (Çarpım)

Şimdi Zincir Kuralını uygulayıp A ve B parçalarını çarpalım. Sihri izle:

$$\frac{\partial Loss}{\partial Z_2} = \text{Adım A} \cdot \text{Adım B}$$

$$\frac{\partial Loss}{\partial Z_2} = \left(\frac{p - y}{p(1 - p)}\right) \cdot (p(1 - p))$$

Gördün mü? Paydadaki ile Sigmoid türevinden gelen birbirini **YOK EDER.**

Geriye sadece şu kalır:

$$\frac{\partial Loss}{\partial Z_2} = p - y$$

İşte kodda `dZ2 = p - Y` yazmamızın bilimsel ispatı budur.

2. Ağırlıkların Türevi (dW) Neden `X.T * Hata`?

Şimdi hatayı (dZ) bulduk. Peki bu hata ağırlıkları (W) nasıl etkiler?

Denklemimiz: Z = X . W + b

Zincir kuralı:

$$\frac{\partial Loss}{\partial W} = \frac{\partial Loss}{\partial Z} \cdot \frac{\partial Z}{\partial W}$$

- Birinci kısım ($\frac{\partial Loss}{\partial Z}$) zaten yukarıda bulduğumuz hata (dZ).
- İkinci kısım ($\frac{\partial Z}{\partial W}$):
  - $Z = x \cdot w$ denkleminde w'ye göre türev alırsan geriye x kalır.

Yani:

$$dW = x \cdot dZ$$

Matris Boyutları İçin Transpose:

- matrisi: `(Örnek Sayısı, Girdi)`
- matrisi: `(Örnek Sayısı, Çıktı)`
- Bizim istediğimiz W: `(Girdi, Çıktı)`

Bu boyutları eşleştirmek için X'i yan yatırırız (X^t):

$$dW = X^T \cdot dZ$$

3. ReLU Türevi Neden `(Z > 0)`?

Gizli katmanda aktivasyon olarak ReLU kullandık: $A = ReLU(Z)$. Fonksiyon şuydu: `Maksimum(0, Z)`

Bunun türevi (eğimi) parçalıdır:

1. **Eğer Z > 0 ise:** Fonksiyon $y = x$ gibidir. Eğimi **1**'dir.
2. **Eğer Z <= 0 ise:** Fonksiyon düz çizgi (0) gibidir. Eğimi **0**'dır.

Kodda `dZ1 = dA1 * (Z1 > 0)` satırı şunu yapar:

- `(Z1 > 0)` ifadesi, `Z1` pozitifse `1`, negatifse `0` üreten bir maske oluşturur.
- Bu maskeyi hatayla çarparız.
- Sonuç: Pozitif nöronlar hatayı geçirir (1 ile çarpılır), negatif nöronlar hatayı öldürür (0 ile çarpılır).

Hepsini tek bir seferde formül yığını olarak verirsem kafa karıştırır. O yüzden senin istediğin gibi **Zincir Kuralını (Chain Rule)** en dıştan en içe doğru, bir soğan soyar gibi açacağız.

Görselleştirmek için akışımız şu (Sondan başa):

$$Loss \leftarrow p \leftarrow Z_2 \leftarrow A_1 \leftarrow Z_1 \leftarrow W_1$$

1. Bölüm: (Çıktı Katmanı Ağırlıkları)

Burada hedefimiz: **Loss fonksiyonunun W2'ye göre değişimi.**

Tam Zincir Formülü

$$\frac{\partial Loss}{\partial W_2} = \underbrace{\frac{\partial Loss}{\partial p} \cdot \frac{\partial p}{\partial Z_2}}_{\text{Adım 1: Hata}} \cdot \underbrace{\frac{\partial Z_2}{\partial W_2}}_{\text{Adım 2: Girdi}}$$

Gördüğün gibi 2 ana parça var. Hadi parçalayalım:

Adım 1: Hatayı Bul (dZ2)

İlk iki terim ($(\frac{\partial Loss}{\partial p} \cdot \frac{\partial p}{\partial Z_2})$), az önceki cevabımda ispatladığımız gibi birbirini sadeleştirir ve şu kalır:

$$dZ_2 = p - Y$$

Adım 2: Ağırlığın Etkisini Bul

Denklemimiz neydi? Z2 = A1 . W2 + b2 Burada W2'ye göre türev alırsak geriye ne kalır?

- Cevap: w2 'nin katsayısı olan (Gizli katmandan gelen aktivasyon).

Sonuç (Birleştirme)

Matematiksel olarak: Hata (dZ2) Girdi (A1). Ancak Matris dünyasında boyutların uyuşması için Girdiyi yan çeviririz (A1.T):

$$dW_2 = A_1^T \cdot dZ_2$$

2. Bölüm: (Gizli Katman Ağırlıkları) - ASIL OLAY BURADA! 🚨

Burada hedefimiz: **Loss fonksiyonunun taaa en baştaki W1'e göre değişimi.** Yolumuz çok uzun. Loss'tan çıkıp W1'e gitmek için 4 kapıdan geçmemiz lazım.

🔗 **Tam Zincir Formülü**

$$\frac{\partial Loss}{\partial W_1} = \underbrace{\frac{\partial Loss}{\partial Z_2}}_{\text{1. Çıkış Hatası}} \cdot \underbrace{\frac{\partial Z_2}{\partial A_1}}_{\text{2. Köprüyü Geç}} \cdot \underbrace{\frac{\partial A_1}{\partial Z_1}}_{\text{3. ReLU Kapısı}} \cdot \underbrace{\frac{\partial Z_1}{\partial W_1}}_{\text{4. Giriş Verisi}}$$

Hadi bu 4 terimi tek tek hesaplayıp çarpalım.

Parça 1: Çıkıştaki Hata (dZ2)

Bunu zaten bulmuştuk. Zincirin başı burasıdır.

$$dZ_2 = p - Y$$

Parça 2: Hatayı Geriye Taşı (üzerinden)

Formülümüz: $Z_2 = A_1 \cdot W_2$ Şimdi sorumuz şu: "**Z_2'deki hata, A_1'i (önceki katmanı) nasıl etkiler?**"

- Yani Z_2'nin A_1'e göre türevini alıyoruz.
- Geriye katsayı olarak W_2 kalır.

Matris İşlemi: Hatayı (dZ_2), ağırlıklarla (W_2) çarparak geriye atıyoruz. Boyut uydurmak için W_2'yi yan çeviriyoruz.

$$dA_1 = dZ_2 \cdot W_2^T$$

(Şu an hatayı Gizli Katmanın kapısına kadar getirdik).

Parça 3: Aktivasyon Türevi (ReLU)

Formülümüz: $A_1 = \mathrm{ReLU}(Z_1)$ Sorumuz: "**A_1'deki değişim, Z_1'i nasıl etkiler?**"

- ReLU'nun türevi basittir: Z_1 pozitifse 1, negatifse 0.

Matris İşlemi: Elimizdeki hatayı (dA_1), türev maskesiyle çarpıyoruz.

$$dZ_1 = dA_1 \cdot (Z_1 > 0)$$

(Artık hatayı ReLU'nun içinden geçirdik, Z_1'e ulaştık).

Parça 4: Giriş Verisi (X)

Formülümüz: $Z_1 = X \cdot W_1$ Son sorumuz: **"Z_1'deki hata, W_1'i nasıl etkiler?"**

- W_1'e göre türev alırsak, geriye katsayı olarak X (Girdi verisi) kalır.

Matris İşlemi: Hata (dZ_1) ile Girdiyi (X) çarp. Boyut için X'i yan çevir.

$$dW_1 = X^T \cdot dZ_1$$

🧩 Büyük Resmi Birleştirelim

Şimdi o uzun kod satırlarının aslında bu zincirin halkaları olduğunu görebilirsin:

1. **Hata oluştu:** `dZ2 = p - Y`
2. **Hata W2 üzerinden geriye aktı:** `dA1 = np.dot(dZ2, W2.T)`
3. **Hata ReLU'dan süzüldü:** `dZ1 = dA1 * (Z1 > 0)`
4. **Hata W1'e ulaştı (X ile çarpıldı):** `dW1 = np.dot(X.T, dZ1)`