Tuğba ULUSOY G171210017

1) a) istenen: $T = \{[1], [2], [3], [5]\}$, hesaplanan: $Y = \{[0], [2], [1], [3]\}$

$$MSE(loss) = \frac{1}{m} \sum_{i=1}^{m} (t_i - y_i)^2 \qquad m = 4 \rightarrow \text{veri noktalar sayısı}$$

$$= \frac{1}{4}\left[(1-0)^2 + (2-2)^2 + (3-1)^2 + (5-3)^2\right]$$

$$= \frac{1}{4}\left[1 + 0 + 4 + 4\right] = \frac{9}{4} = \boxed{2.25}$$

b) istenen: $T\left\{ \begin{array}{c} A \\ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \end{array}, \begin{array}{c} B \\ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{array}, \begin{array}{c} C \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{array} \right\}$, hesaplanan: $Y = \left\{ \begin{array}{c} A \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{array}, \begin{array}{c} B \\ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \end{array}, \begin{array}{c} C \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{array} \right\}$

$$\text{Categorical cross entropy} = -\sum_{i=1}^{c} t_i \log(y_i)$$, $c = 3 \rightarrow$ kategori sayısı ve A, B, C $\rightarrow$ eğitim örnekler olsun.

A için: $T\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $Y\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$$-0.\log_{10} 1 - 1\log_{10}(0+\alpha) - 0.\text{Log}_{10}(0+\alpha)$$

Log 0 tanımsız olduğu için $\alpha \in \mathbb{R}$ keyfi küçük değer eklememiz lazım. $\alpha = 10^{-14}$ seçersek

$$\Rightarrow \left[0 - 1.\log(0 + 10^{-14}) - 0\right]$$

$$\Rightarrow -\text{Log}_{10} 10^{-14} = \boxed{14}$$

B için: $T = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $Y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$$\left[-0.\text{Log}(0 + 10^{-14}) - 0.\text{Log} 1 - 1.\text{Log}(0 + 10^{-14})\right]$$

$$\Rightarrow 0 - 0 + 14 = \boxed{14}$$

C için: $T = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $Y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$$\left[-1.\text{Log} 1 - 0.\text{Log}(0 + 10^{-14}) - 0.\text{Log}(0 + 10^{-14})\right]$$

$$\Rightarrow \boxed{0}$$

Sonuç olarak;

A, B, C categorical cross entropy loss $= \left[14, 14, 0\right]$

Tuğba ULUSOY G171210017

2) Eğer modelimiz, eğitim için kullandığımız veri setimiz üzerinde gereğinden fazla çalışıp ezber yapmaya başlamışsa ya da eğitim setimiz setimiz tekdüze ise overfitting olma riski büyük demektir.

Overfitting problemi olan modellerde yüksek varyans, düşük bias durumu görülmektedir. Bu genellikle model çok karmaşık olduğunda (yani gözlem sayısına kıyasla çok fazla özellik / değişken varsa) gerçekleşir.

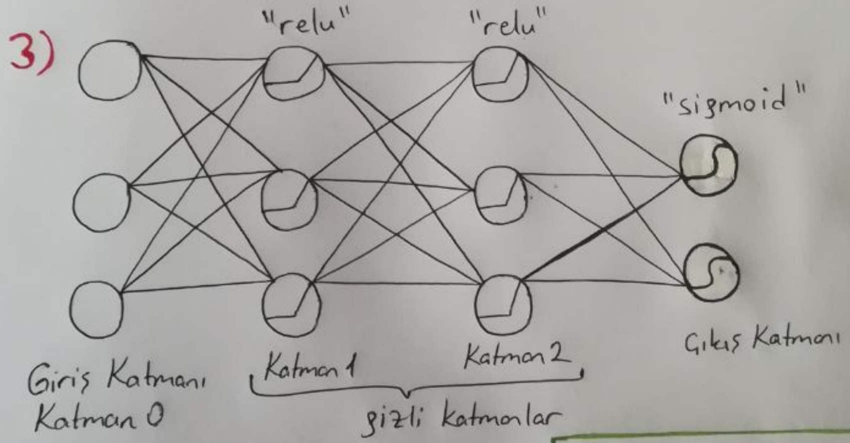
Overfitting problemi aşağıdaki yöntemler uygulanarak çözülebilmektedir;

* Öz nitelik sayısını azaltmak: Birbirleriyle yüksek korelasyonlu olan kolonlar silinebilir ya da faktör analizi gibi yöntemlerle bu değişkenlerden tek bir değişken oluşturulabilir.

* Daha fazla veri eklemek: Eğer eğitim seti tek düze ise daha fazla veri ekleyerek veri çeşitliliği arttırılır.

* Regularization (Düzenleme): Modelin karmaşıklığını azaltmak için kullanılan bir tekniktir. Bunu kayıp fonksiyonunu cezalandırarak yapar. Yani modelde ağırlığı yüksek olan değişkenlerin ağırlığını azaltarak bu değişkenlerin etki oranını azaltır. Değişkenlerin ağırlığını azaltmak için regularization değerini arttırmak gerekmektedir. En popüler regularization metotları Lasso ve Ridge teknikleridir.

3)



⇒ Eğitilebilir parametre sayısı her katman için;
(katmandaki nöron sayısı X (önceki katmanın nöron sayısı + 1))

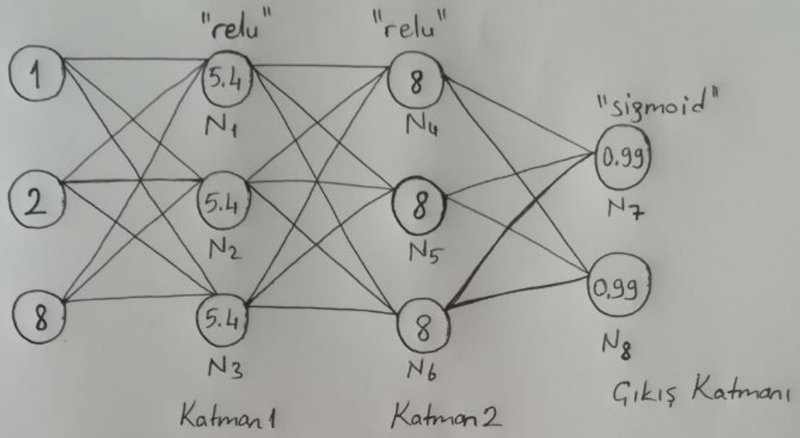
Katman 0 : Eğitilemez
Katman 1 : $3.(3+1) = 3.4 = \boxed{12}$
Katman 2 : $3.(3+1) = 3.4 = \boxed{12}$
Çıkış Katmanı : $2(3+1) = 2.4 = \boxed{8}$

Toplam eğitilebilir parametreler ;

$12 + 12 + 8 = \boxed{32}$

Tuğba ULUSOY G171210017

4) giriş $X = \begin{bmatrix} 0+1 \\ 1+1 \\ 7+1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix}$



$N_i = A(V)$, $V = \sum_{i=0}^{n} (w_i X_i + b)$

A: Aktivasyon fonksiyon (relu, sigmoid vb.)

x_i : giriş değerler

w_i : ağırlıklar

b : bias değeri

N_i : nöron

"relu" olursa $A = max(0, V)$

"Sigmoid" olursa $A = \frac{1}{1+e^{-V}}$

Böylelikle;

Katman1 için; A:"relu" → max(0,V)

$$V = \sum_{i=1}^{n} w_i x_i + b \Rightarrow V = 0.5(1) + 0.5(2) + 0.5(8) - 0.1 = \boxed{5.4}$$

$\boxed{N_1 = max(0, 5.4) = 5.4}$ $\boxed{N_1 = N_2 = N_3}$ (Aynı ağırlıklara sahip oldukları için)

Katman 2 için; A:"relu" → max(0,V)

$V = 0.5(5.4) + 0.5(5.4) + 0.5(5.4) - 0.1$ ⟶ $\boxed{V = 8}$

$\boxed{N_4 = max(0, 8) = 8}$ $\boxed{N_4 = N_5 = N_6} \boxed{= 8}$

Çıkış Katman için; A:"sigmoid" ⟶ $\frac{1}{1+e^{-V}}$

$N_7 = \frac{1}{1+e^{-11.9}} = 0.99$

$V = 0.5(8) + 0.5(8) + 0.5(8) - 0.1$

$\boxed{V = 11.9}$

$\boxed{N_7 = N_8 =} \boxed{0.99}$