

Subject:

Year: Month: Date:

810100251 / محمد عبد

بعض مسائل

$$a_j = \sum_i (w_{ij} \times x_i), y_j = F(a_j) = \frac{1}{1 + e^{-a_j}}$$

① الت

$$h_1 \rightarrow a_3 = 0.25 + 0.1 \times 0.5 + 0.3 \times 0.15 = 0.3, o(h_1) = F(a_3) = \frac{1}{1 + e^{-0.3}} \approx 0.57$$

$$h_2 \rightarrow a_4 = 0.35 + 0.4 \times 0.15 + 0.2 \times 0.05 = 0.42, o(h_2) = F(a_4) = \frac{1}{1 + e^{-0.42}} \approx 0.6$$

$$o_1 \rightarrow a_5 = 0.45 + 0.5 \times 0.57 + 0.6 \times 0.6 = 1.09, o(o_1) = F(a_5) = \frac{1}{1 + e^{-1.09}} \approx 0.73$$

$$o_2 \rightarrow a_6 = 0.55 + 0.6 \times 0.57 + 0.8 \times 0.6 = 1.37, o(o_2) = F(a_6) = \frac{1}{1 + e^{-1.37}} \approx 0.79$$

$$E = \frac{1}{2} \left[ (\underbrace{0.73}_{0.39} - \underbrace{0.1}_{0.01})^2 + (\underbrace{0.79}_{0.101} - \underbrace{0.9}_{0.01})^2 \right] = 0.2$$

$$\Delta w_{ij} = \eta \delta_j o_i$$

ج 1) اده سة فريد هال سة فريد

$$\delta_j = o_j (1 - o_j) (t_j - o_j) \quad j = \text{output unit}$$

$$\delta_j = o_j (1 - o_j) \sum_k \delta_k w_{kj} \quad j = \text{hidden unit}$$

$$\delta_{o_1} = 0.73 (1 - 0.73) (0.1 - 0.73) = -0.12$$

حال دايح

$$\Delta w_5 = \eta \times 0.57 \times -0.12 = 0.5 \times 0.57 \times -0.12 = -0.03$$

learning rate

$$w_5 (\text{new}) = w_5 (\text{old}) + \Delta w_5 = 0.5 + (-0.03) = 0.47$$



Epoch  
000,855

Learning rate  
0.03

Activation  
ReLU

Regularization  
None

Regularization rate  
0

Problem type  
Classification

### FEATURES

Which properties do you want to feed in?

- $X_1$
- $X_2$
- $X_1^2$
- $X_2^2$
- $X_1 X_2$
- $\sin(X_1)$
- $\sin(X_2)$

+ - 1 HIDDEN LAYER

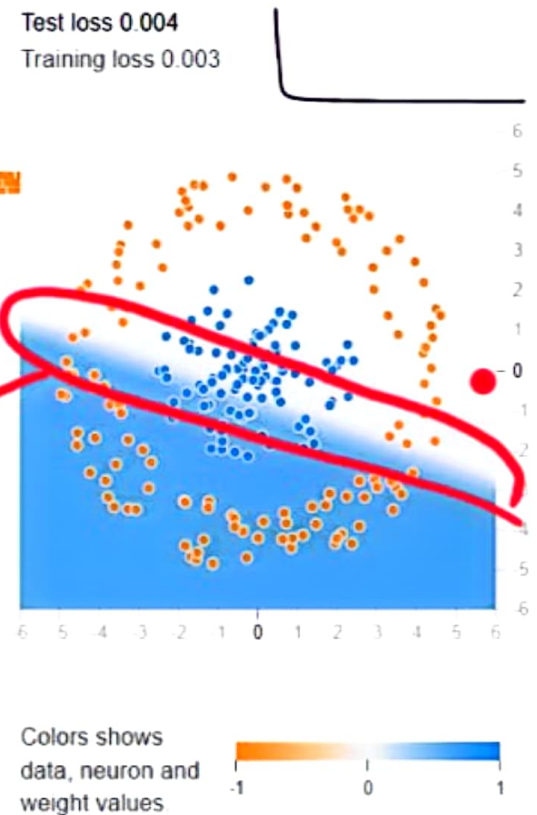
+ -  
3 neurons

This is the output from one **neuron**  
Hover to see it larger

### OUTPUT

Test loss 0.004  
Training loss 0.003

decision boundary



000,855

0.03

ReLU

None

0

## Classification

+ - 1 HIDDEN LAYER

+

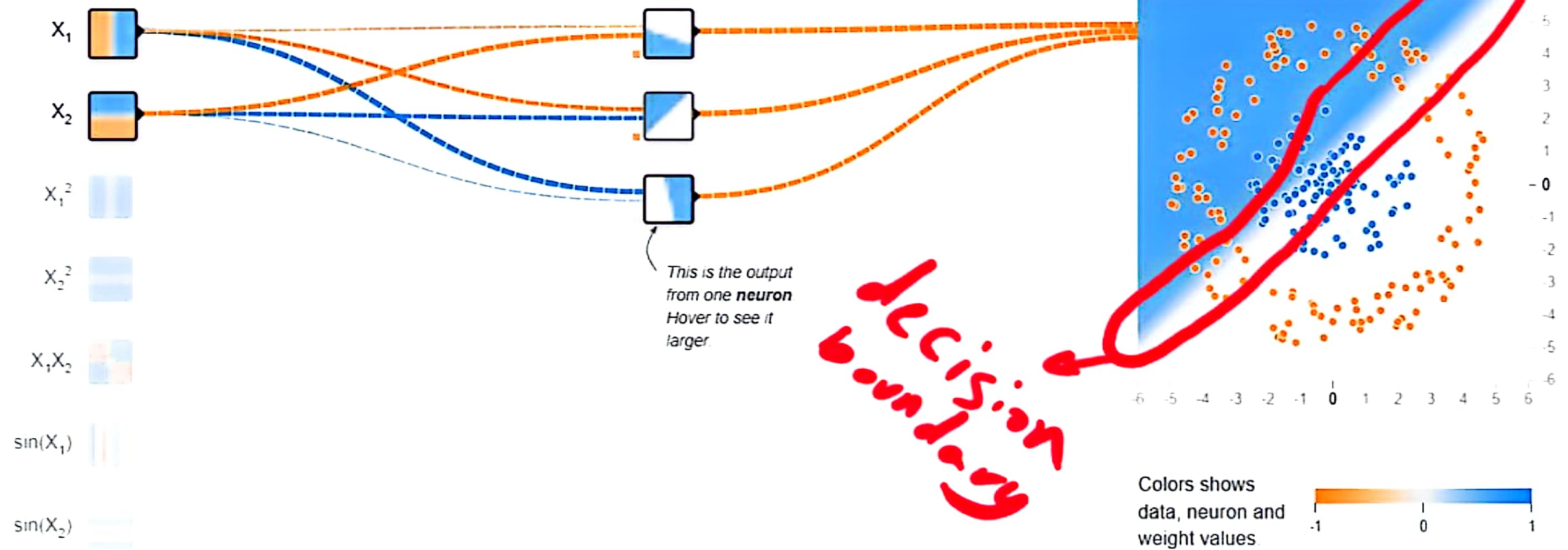
-

3 neurons

## OUTPUT

Test loss 0.004

Training loss 0.003





Epoch 000,855      Learning rate 0.03      Activation ReLU      Regularization None      Regularization rate 0      Problem type Classification

## FEATURES

Which properties do you want to feed in?

- $X_1$
- $X_2$
- $X_1^2$
- $X_2^2$
- $X_1X_2$
- $\sin(X_1)$
- $\sin(X_2)$

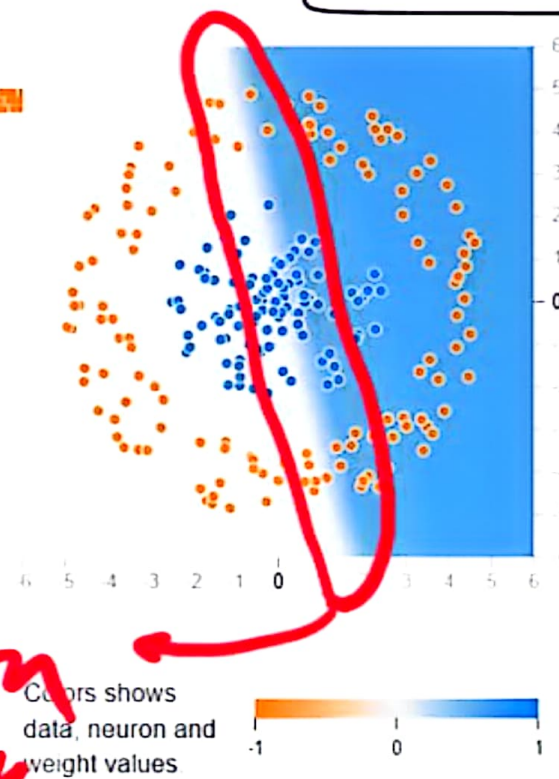
+ - 1 HIDDEN LAYER

+ -  
3 neurons

This is the output from one neuron  
Hover to see it larger

## OUTPUT

Test loss 0.004  
Training loss 0.003



decision boundary

الف) چون  $n_2$  در استای مخورک هست و  $n_{22}$  در استای مخورک هست به وزن نیازیم  
hidden layer و مدل دایره به شکل  $(n_1 - n_2)^2 + (n_1 - n_2)^2 - \alpha = 0$  است می توانیم تعیین کنیم.

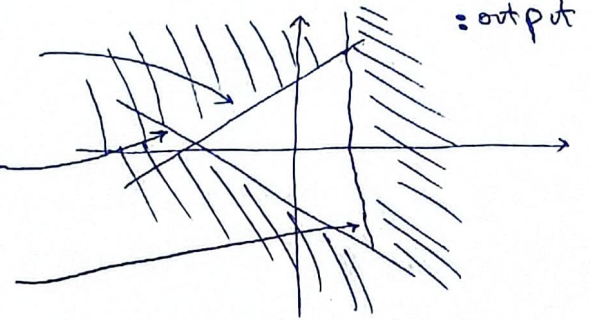
بایاس

ب) عدد در epoch=10

نورون اول =  $2.8 n_1 + (-1.91) n_2 + -7.31$

نورون دوم =  $4.7 n_1 + -1.6 n_2 + 4.4$

نورون سوم =  $-1.6 n_1 - 2.8 n_2 - 7.85$



ب) مثلاً نقطه  $(4, 1)$  در نورون اول و سوم پس از فرج از تابع فعال سازی می شود.  
ایجاد نورون دوم خروجی منفی weight آن هم منفی است، پس در خروجی باید مثبت باشد دهانی  
که منفی طبقه شده پس بایاس باید مثبت باشد.

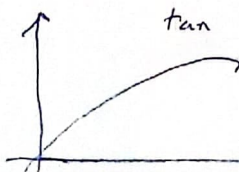
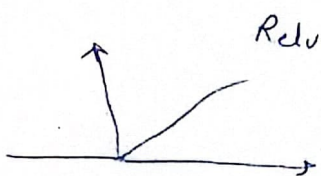
$Relu \left( (-3, 3) \times (\text{نورون دوم}) + (4, 6) \times (\text{نورون دوم}) + (-3, 4) \times (\text{نورون اول}) + \text{بایاس} \right)$

$(-0.6 + (-1.91) + 5.28) n_1 + (6.46 + (-7.36) + 7.24) n_2$   
 $\underbrace{\quad}_{2.16} \quad \underbrace{\quad}_{8.3}$   
 $7.3 n_1 \quad 8.3 n_2$

$(1.05 + 20.24 + 2.8) < 24.9$

بایاس لایه آخر

$7.3 n_1 + 8.3 n_2 + 24.9 + \alpha$



ب) با توجه به

در حالت Relu تغییر داریم چون فراد شکلی خاص تندی داریم اما در حالت تانجس چون از تابع فعال سازی غیر خطی است دهانی

برای رگرسیون هم همین طور و حالت دلخواه ای داریم  
بنویسیم

شبکه های عمیق پیچیده

- (1) جلوگیری از overfitting: کاهش مقدار پارامترها: لایه های پیچیده با اتصال های خیلی قوی و مقدار پارامترهای کمتر و کمتر که باعث کاهش پیچیدگی مدل و overfitting کمتری می شود.
- (2) حفظ استقلال ویژگی های مکانی: لایه های پیچیده می توانند ویژگی های مکانی داشته باشند و باعث وابستگی ها را به فرم شناختی می کنند و باعث می شود لایه های بالاتر اتصال این ویژگی های مکانی را از بین ببرند.
- (3) تقیید ناچیزی نسبت به انتقال و تغییرات جزئی: لایه های پیچیده و ادما (شبکه) نسبت به تغییرات جزئی و انتقال های مکانی مقاوم تر می کنند که باعث دقت در شناسایی اشیاء می شود.
- (4) کاهش مصرف حافظه و افزایش کارایی محاسباتی: لایه های پیچیده با پردازش همزمان استفاده از ویژگی های کوچک مصرف حافظه را کاهش می دهد و محاسباتی را بهینه می کند.
- (5) weight sharing: در لایه های پیچیده یک فیلتر واحد در هر تصویر اعمال می شود به این معنی که پارامترها به اشتراک می افتد و باعث می شود که پارامترها به اشتراک می افتد و باعث می شود که پارامترها به اشتراک می افتد و باعث می شود که پارامترها به اشتراک می افتد.



$$f: \mathbb{R}^{1 \times 3} \rightarrow [f_1, f_2, f_3] \rightarrow \begin{cases} f_1 + 4f_2 + 0f_3 = -2 \rightarrow f_1 + 4f_2 = -2 \\ 4f_1 + 0f_2 + (-2)f_3 = -2 \rightarrow 4f_1 - 2f_3 = -2 \\ 0f_1 + (-62)f_2 + 3f_3 = 11 \rightarrow -2f_2 + 3f_3 = 11 \end{cases}$$

$$\rightarrow \begin{cases} f_2 = -\frac{10}{13} \\ f_3 = \frac{41}{13} \\ f_1 = \frac{14}{13} \end{cases} \rightarrow f = \left[ \frac{14}{13}, -\frac{10}{13}, \frac{41}{13} \right]$$