

## بهینه‌سازی

### سوال اول

برای راحتی اعداد تا دو رقم اعشار گرد شده‌اند.  
(الف)

$$h_1 = \sigma(0.05 \times 0.1 + 0.15 \times 0.3 + 0.25) = 0.56$$

$$h_2 = \sigma(0.05 \times 0.2 + 0.15 \times 0.4 + 0.35) = 0.6$$

$$O_1 = \sigma(0.56 \times 0.5 + 0.6 \times 0.7 + 0.45) = 0.76$$

$$O_2 = \sigma(0.56 \times 0.6 + 0.6 \times 0.8 + 0.55) = 0.8$$

(ب)

$$E = \frac{1}{2} [(0.1 - 0.76)^2 + (0.9 - 0.8)^2] = 0.22$$

(ج)

$$\frac{\partial E}{\partial W_5} = \frac{\partial E}{\partial O_1} \times \frac{\partial O_1}{\partial W_5}$$

$$E = \frac{1}{2} [(t_1 - o_1)^2 + (t_2 - o_2)^2] \rightarrow \frac{\partial E}{\partial O_1} = -(t_1 - O_1)$$

$$O_1 = \sigma(h_1 \times W_5 + h_2 \times W_1 + b_3) \rightarrow \frac{\partial O_1}{\partial W_5} = h_1 \times \frac{\partial \sigma(h_1 \times W_5 + h_2 \times W_1 + b_3)}{\partial W_5}$$

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x) \times (1 - \sigma(x))$$

$$\frac{\partial \sigma(h_1 \times W_5 + h_2 \times W_1 + b_3)}{\partial W_5} = h_1 \times \sigma(h_1 \times W_5 + h_2 \times W_1 + b_3) \times (1 - \sigma(h_1 \times W_5 + h_2 \times W_1 + b_3))$$

$$\frac{\partial E}{\partial W_5} = -(t_1 - O_1) \times h_1 \times (\sigma(h_1 \times W_5 + h_2 \times W_1 + b_3) \times (1 - \sigma(h_1 \times W_5 + h_2 \times W_1 + b_3)))$$

$$\frac{\partial E}{\partial W_5} = -(0.1 - 0.76) \times 0.57 \times (0.76 \times (1 - 0.76)) = 0.06$$

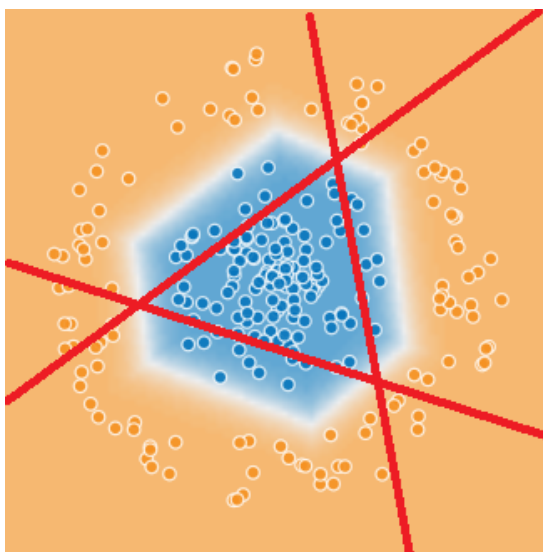
$$W_5 = W_5 - \alpha \frac{\partial E}{\partial W_5} \rightarrow W_5 = 0.5 - 0.06 \times 0.5 = 0.47$$

## شبکه های عصبی

### سوال اول

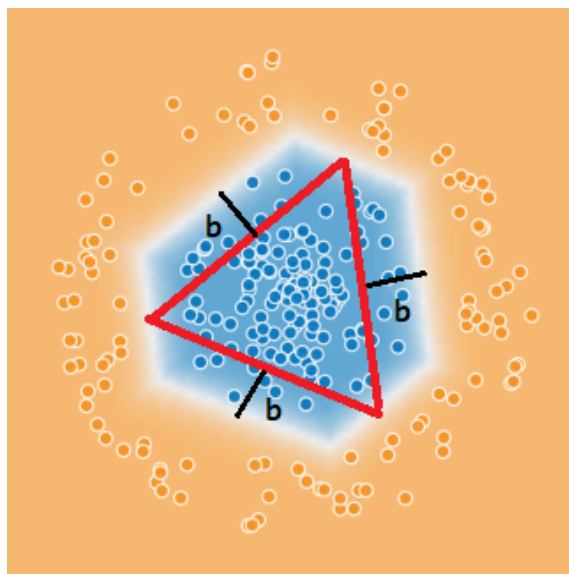
الف) می‌دانیم که دو فیچر  $X_1^2$  و  $X_2^2$  مستقیماً به لایه خروجی می‌روند در نتیجه مقدار ورودی تابع فعال‌ساز این لایه برابر  $W_1 \times X_1^2 + W_2 \times X_2^2 + b$  می‌باشد که فرمول بیضی است. اگر  $W_1 \approx W_2$  آنگاه مرز تصمیم یک دایره می‌شود، که خود شبکه این را آموزش می‌بیند. (می‌توانید به وزن‌ها پس از آموزش دقت کنید).

ب) ابتدا معادله خروجی هر نورون را به دست می‌آوریم. ورودی تابع فعال‌ساز ReLU برابر  $W_1 \times X_1 + W_2 \times X_2 + b$  می‌باشد که یک معادله خط است. یک سمت این خط مقادیر مثبت و سمت دیگر مقادیر منفی می‌باشند. با توجه به تابع ReLU، مرز تصمیم یک نورون، یک خط می‌باشد که یک طرف آن مقادیر مثبت (مکان‌هایی که با جایگذاری  $X_1$  و  $X_2$  آن‌ها در معادله بالا، مقدار بزرگ‌تر از صفر تولید می‌کنند) و یک طرف دیگر آن **کاملاً صفر** است. برای هر سه نورون این مرز را ترسیم می‌کنیم:



با توجه به معادله‌ها، در ناحیه مثلث ایجاد شده، هر سه نورون مقدار صفر را برمیگردانند. از طرفی می‌دانیم مرز تصمیم لایه خروجی دارای فرمول  $W_1 \times O_1 + W_2 \times O_2 + W_3 \times O_3 + b$  می‌باشد. در نتیجه برای این ناحیه چون هر سه مقدار  $O_1$ ،  $O_2$  و  $O_3$  صفر می‌باشند، تنها بایاس این لایه، تصمیم نهایی را می‌گیرد که با توجه به مثبت بودن این ناحیه، بایاس لایه خروجی باید مثبت باشد.

حال که فهمیدیم بایاس لایه خروجی مثبت است، می‌توان مرز تصمیم را این گونه تعبیر کرد:



پ) تابع‌های  $\tanh$  و  $\text{sigmoid}$  نسبت به  $\text{ReLU}$  دارای smoothness بیشتری می‌باشند، در نتیجه مرز تصمیم  $\text{ReLU}$  نسبت به این دو تیزتر است.

## شبکه های عصبی پیچشی

### سوال اول

هدف استفاده از لایه های کانولوشنی و پولینگ، یادگیری فیچرهای عکس است. اگر از این لایه ها استفاده نکنیم و پیکسل های یک عکس را مستقیماً به یک شبکه عصبی دهیم، علاوه بر محاسبه زیاد و سنگین، مدل پیکسل های عکس ها را حفظ می کند و نمی تواند الگو و فیچر از عکس استخراج کند، (زیرا یک پیکسل به تنهایی اطلاعات خاصی ندارد ولی معمولاً می توان اطلاعات معناداری از چندین پیکسل کنار هم استخراج کرد که این به علت هم بستگی مکانی پیکسل ها در عکس می باشد). به همین علت مدل اختلاف دقت قابل توجهی بر روی داده های ترین و تست خواهد داشت و اورفیت خواهد کرد.

### سوال دوم

فیلتر  $f$  را مانند زیر در نظر می گیریم:

$$f = [h_1, h_2, h_3]$$

$$h_1 + 4h_2 = -2$$

$$4h_1 - 2h_3 = -2$$

$$-2h_2 + 3h_3 = 11$$

$$h_1 = \frac{14}{13}, h_2 = -\frac{10}{13}, h_3 = \frac{41}{13}$$