

سوال اول:

الف) آکسون: زائده منفرد و بلند، وظیفه آن فرستنده اصلی نورون است.

دندريت: از زوائد نورون، ياخته‌های خونی معمولاً بیش از یک عدد، کوتاه و منشعب، ختم به انشعابات ظریف و متعدد در انتها است. این بخش گیرنده اصلی نورون است.

هسته: اطلاعاتی که از دندريت‌ها دریافت شده است توسط هسته پردازش می‌شود.

ب) ارتباط بین دندريت یک نورون و آکسون نورون‌های دیگر، سیناپس نام دارد. وظیفه آن انتقال پالس الکتریکی میان دو نورون می‌باشد.

پ) دندريت = ورودی ها / هسته = وزن ها / آکسون = خروجی

ت) در شبکه های مصنوعی اگر یک سلول آسیب ببیند بقیه سلول‌ها می‌توانند نبود آن را جبران کرده، و نیز در بازسازی آن سهیم باشند.

شبکه های مصنوعی قادر به یادگیری‌اند.

شبکه های مصنوعی لایه لایه می‌باشند و نورون های هر لایه با یکدیگر ارتباط ندارند.

در شبکه های مصنوعی ارتباط میان نورون ها قابل تغییر می‌باشند.

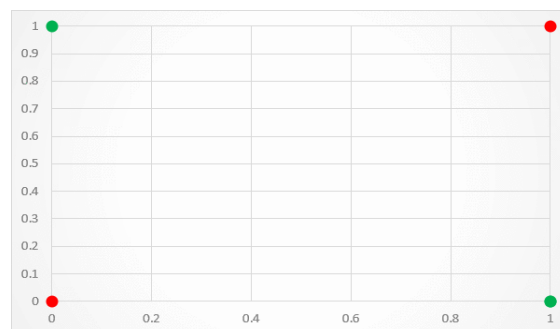
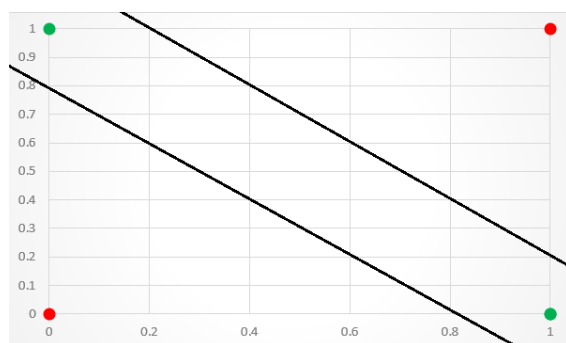
سوال دوم:

الف) زیرا میتوان با استفاده از آن، مقداری که می‌خواهیم نورون بعد از آن فعال شود را تعیین کنیم و باعث افزایش انعطاف‌پذیری تابع فعال‌سازی می‌شود.

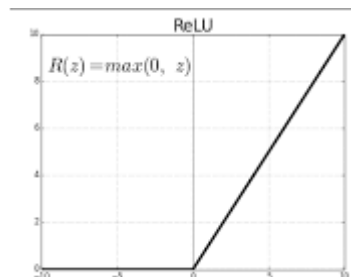
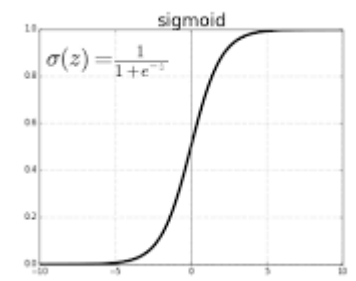
ب) خروجی تابع XOR فقط ۴ نقطه در فضا می‌باشند و یک نورون می‌تواند تنها یک خط راست رسم کند بنابراین به تنهایی نمی‌تواند خروجی‌ها را از هم تفکیک کند.

پ) خیر، چون با داشتن نورون‌های بیشتر و لایه‌های بیشتر می‌توان منحنی‌های دقیق‌تری برای تفکیک خروجی‌های XOR کشید ولی در این مثال ما نیاز به دو خط موازی داریم.

ت) بله، زیرا با استفاده از این توابع و همچنین چند لایه‌ای بودن شبکه می‌توان خروجی‌های XOR را تفکیک کرد.



ث) تفاوت اصلی میان این دو تابع این است که بعد از صفر تابع sigmoid به یک میل می‌کند ولی تابع ReLU به بی‌نهایت می‌رود. تابع ReLU کاربردی‌تر است (به دلیل اسپارس بودن) و سرعت یادگیری شبکه با استفاده از این تابع بیشتر می‌شود.



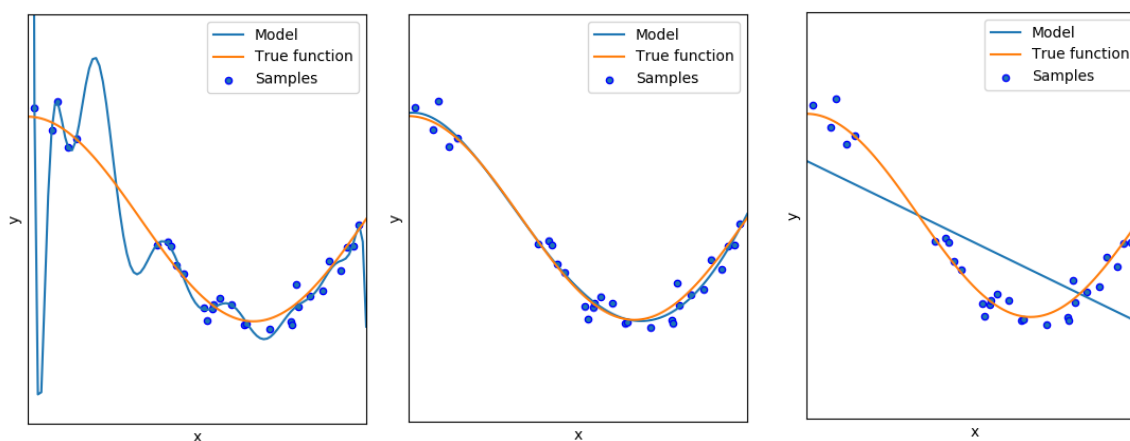
سوال سوم:

الف) در بایاس بالا، **Underfitting** اتفاق می‌افتد به این معنی که ماشین حتی برای نمونه‌های قبلا دیده شده نیز پاسخ دقیق نمی‌دهد. در واریانس بالا، **Overfitting** اتفاق افتاده است بدین معنی که ماشین برای داده‌ها دیده شده پاسخ دقیق می‌دهد اما برای نمونه‌های جدید و از قبل دیده نشده پاسخ‌های پرتی می‌دهد.

ب) اگر شبکه ما در برای نمونه‌های دیده شده جواب دقیقی نمی‌دهد، بایاس بالا رخ داده است. اگر شبکه برای داده‌های دیده شده به خوبی پاسخ می‌دهد ولی برای داده‌های دیده نشده پاسخ پرتی می‌دهد، آنگاه واریانس بالا اتفاق افتاده است.

پ) کافی است عمل آموزش (Epoch) و تغییر وزن‌ها را بیشتر انجام دهیم.

ت) در تکنیک **dropout** در پروسه آموزش به صورت رندوم بعضی از نورون‌ها را ایگنور می‌کنیم. در تکنیک **Regularization** دو روش **LASSO** و **Ridge** وجود دارند که هر دو با استفاده از **Regression** باعث جلوگیری از واریانس بالا می‌شوند.



واریانس بالا

مدل مناسب

بایاس بالا

سوال چهارم:

(الف)

$$\frac{\partial Cost}{\partial w_1} = \frac{\partial Cost}{\partial A_0} \times \frac{\partial A_0}{\partial Z_{A_0}} \times \frac{\partial Z_{A_0}}{\partial w_1} = \frac{\partial Cost}{\partial A_0} \times \sigma'(Z_{A_0}) \times X_1$$

$$\frac{\partial Cost}{\partial A_0} = \frac{\partial Cost}{\partial B_0} \times \frac{\partial B_0}{\partial Z_{B_0}} \times \frac{\partial Z_{B_0}}{\partial A_0} = \frac{\partial Cost}{\partial B_0} \times \sigma'(Z_{B_0}) \times v_0$$

$$\begin{aligned} \frac{\partial Cost}{\partial B_0} &= \frac{\partial Cost}{\partial O_1} \times \frac{\partial O_1}{\partial Z_{O_1}} \times \frac{\partial Z_{O_1}}{\partial B_0} = \frac{\partial Cost}{\partial B_0} \times \sigma'(Z_{O_1}) \times t_0 \\ &= (2O_1 - 2y_t) \times \sigma'(Z_{O_1}) \times t_0 \end{aligned}$$

(ب)

$$\begin{aligned} A_0 &= \sigma(b_0 + w_0X_0 + w_1X_1 + w_2X_2) = \sigma(0.5 + 0.4 \times 1 + 0.3 \times 1 + 0.2 \times 1) \\ &= \sigma(1.4) = 0.802 \end{aligned}$$

$$\begin{aligned} A_1 &= \sigma(b_1 + u_0X_0 + u_1X_1 + u_2X_2) = \sigma(0.2 + 0.3 \times 1 + 0.4 \times 1 + 0.5 \times 1) \\ &= \sigma(1.4) = 0.802 \end{aligned}$$

$$\begin{aligned} B_0 &= \sigma(b_2 + v_0A_0 + v_1A_1) = \sigma(0.5 + 0.6 \times 0.802 + 0.7 \times 0.802) = \sigma(1.542) \\ &= 0.823 \end{aligned}$$

$$O_1 = \sigma(b_3 + t_0B_0) = \sigma(0.5 + 0.9 \times 0.823) = \sigma(1.240) = 0.775$$

$$E = (O_1 - y_t)^2 = (0.775 - 0)^2 = 0.6$$

سوال پنجم:

طبق محاسبات قسمت اول سوال ۴ داریم

$$\frac{\partial Cost}{\partial w_1} = (2O_1 - 2y_t) \times \sigma'(Z_{O_1}) \times t_0 \times \sigma'(Z_{B_0}) \times v_0 \times \sigma'(Z_{A_0}) \times X_1$$

و میدانیم که

$$\sigma'(X) = \sigma(X) \times (1 - \sigma(X))$$

پس

$$\frac{\partial Cost}{\partial w_1} = 0.003$$

به همین صورت برای وزن های دیگر هم چون ورودی ها برابرند داریم

$$\frac{\partial Cost}{\partial w_0} = 0.003$$

$$\frac{\partial Cost}{\partial w_2} = 0.003$$

با احتساب نرخ یادگیری 0.1 داریم

$$w_0 = w_0 - 0.1 \times \frac{\frac{\partial Cost}{\partial w_0}}{\|\nabla Cost\|} = 0.34$$

$$w_1 = w_1 - 0.1 \times \frac{\frac{\partial Cost}{\partial w_1}}{\|\nabla Cost\|} = 0.24$$

$$w_2 = w_2 - 0.1 \times \frac{\frac{\partial Cost}{\partial w_2}}{\|\nabla Cost\|} = 0.14$$