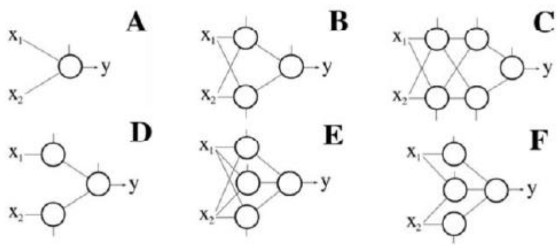
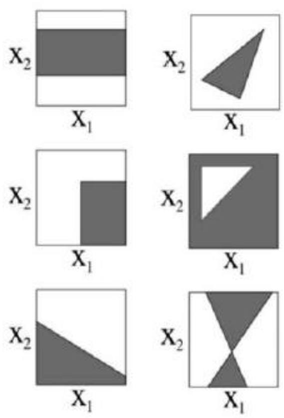
شبکه‌های عصبی پرسپترون چند لایه با تابع فعالسازی پله داده شده در



قسمت بالا از شکل روبرو را در نظر بگیرید. ابتدا مشخص کنید هر کدام از پاسخ‌های داده شده در قسمت پایین از شکل روبرو توسط کدام شبکه تولید شده است و سپس علت را توضیح دهید. فرض کنید که هر کدام از شبکه‌ها فقط یک پاسخ را ایجاد کرده است.



۳- می‌خواهیم مدلی بسازیم که با گرفتن ورودی X، خروجی Y را تخمین بزند. فرض کنید M داریم که روی نمونه‌های bootstrap شده‌ای از پایگاه داده اصلی آموزش داده شده‌اند و دارای بایاس و واریانس یکسان هستند. مدل نهایی به شکل زیر تعریف می‌شود:

$$f_{ensemble}(X) = \frac{1}{M}\sum_{i=1}^{M} f_i(X)$$

الف) بایاس و واریانس مدل $f_{ensemble}(X)$ را بر حسب بایاس و واریانس مدل‌های ضعیف محاسبه کنید. در این قسمت فرض کنید مدل‌های ضعیف از یکدیگر مستقل هستند. تحلیل کنید که افزایش تعداد مدل‌های ضعیف (M) چه تاثیری روی بایاس و واریانس مدل نهایی دارد.

ب) حال فرض کنید مدل‌های ضعیف مستقل نیستند و همبستگی بین هر دو مدل $f_i(X)$ و $f_j(X)$ که $i \neq j$ برابر ρ است. حال مانند قسمت الف، بایاس و واریانس مدل نهایی را محاسبه کرده و تاثیر تعداد مدل ها (M) و وابستگی بین آنها (ρ) را روی این مقادیر بررسی کنید.

③ $f_{ensemble}(X) = \frac{1}{M}\sum_{i=1}^{M} f_i(x)$

الف) $\mathbb{E}(f(x)) = \mathbb{E}\{\frac{1}{M}\sum_{i=1}^{M} f_i(x)\} = \frac{1}{M}\sum_{i=1}^{M} \underbrace{\mathbb{E}(f_i(x))}_{=\bar{f}} = \frac{M}{M}\bar{f} = \bar{f}_i \equiv \mu$

$Var(f(x)) = \mathbb{E}\{(\frac{1}{M}\sum_{i=1}^{M} f_i(x) - \mu)^2\} = \frac{1}{M^2}\left[\mathbb{E}\{\sum_{i=1}^{M} f_i^2(x)\} + M\mu^2 - 2\mu^2\right]$

$= \frac{1}{M^2}\{M(\sigma^2 + \mu^2 - \mu^2)\} = \frac{1}{M}\sigma^2$

ب) $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \longrightarrow \rho = \frac{\mathbb{E}\{f_i(x) f_j(x) - \mu^2\}}{\sigma^2}$

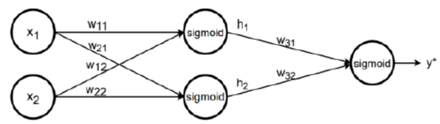
$$\rightarrow \mathbb{E}\{f_i(x) f_j(x)\} = \mu^2 + \rho\sigma^2$$

$$Var(f(x)) = \frac{1}{M^2}\left[M\sigma^2 + M(M-1)(\mu^2 + \rho\sigma^2)\right] - \mu^2 = \rho\sigma^2 + \frac{\sigma^2}{M}(1-\rho)$$

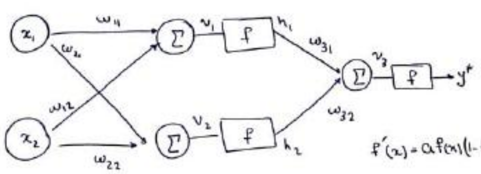
$$\mathbb{E}\left(\frac{1}{M^2}\left(\underbrace{\sum_{i=1}^{M} f_i^2(x)}_{M\sigma^2} + \underbrace{\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} f_i(x) f_j(x)}_{M(M-1)(\mu^2+\rho\sigma^2)} \right) + \underbrace{\mu^2 - 2\mu \sum_{i=1}^{M} f_i(x)}_{-\mu^2} \right)$$

شبکه عصبی زیر را در نظر بگیرید. با استفاده از روش پس انتشارخطا، مشتق عبارت زیر را نسبت به وزن W_{12} بدست اورید (در جواب نهایی می‌بایست از متغیرهایی که در شکل نامگذاری شده است استفاده نمایید). بایاس نرون‌ها در شکل نشان داده نشده است. متغیر y برچسب واقعی است.

$$H(y, y^*) = -[y\log y^* + (1-y)\log(1-y^*)]$$



5



$f'(x) = a f(x)(1-f(x))$

$$\begin{cases} y^* = f(v_3) \\ v_3 = h_1 w_{31} + h_2 w_{32} = f(v_1) w_{31} + f(v_2) w_{32} \\ v_1 = w_{11} x_1 + w_{12} x_2 \\ v_2 = w_{21} x_1 + w_{22} x_2 \\ h_1 = f(v_1) \\ h_2 = f(v_2) \end{cases}$$

$$\frac{\partial H}{\partial w_{12}} = ?$$

$$\frac{\partial H}{\partial w_{12}} = \frac{\partial H}{\partial v_3} \frac{\partial v_3}{\partial w_{12}}$$

$$\frac{\partial v_3}{\partial w_{12}} = \frac{\partial v_3}{\partial h_1}\frac{\partial h_1}{\partial w_{12}} + \frac{\partial v_3}{\partial h_2}\frac{\partial h_2}{\partial w_{12}}$$

$$\frac{\partial h_1}{\partial w_{12}} = \frac{\partial h_1}{\partial v_1}\frac{\partial v_1}{\partial w_{12}} = \frac{\partial h_1}{\partial v_1} x_2 \quad ; \quad \frac{\partial h_1}{\partial v_1} = f'(v_1)$$

$$\frac{\partial h_2}{\partial w_{12}} = \frac{\partial h_2}{\partial v_2}\frac{\partial v_2}{\partial w_{12}} = 0 \quad ; \quad \frac{\partial v_3}{\partial h_1} = w_{31} \; ; \; \frac{\partial v_3}{\partial h_2} = w_{32}$$

$$\rightarrow \frac{\partial H}{\partial w_{12}} = \frac{\partial H}{\partial v_3}\left(w_{31}\, f'(v_1)\, x_2 \right)$$

$$\frac{\partial H}{\partial v_3} = \frac{\partial H}{\partial y^f}\,\frac{\partial y^f}{\partial v_3} = f'(v_3)\left\{ -\frac{y}{y^f} + \frac{1-y}{1-y^f} \right\}$$

$$\rightarrow \frac{\partial H}{\partial v_3} = a f(v_3)(1-f(v_3))\left\{ -\frac{y}{f(v_3)} + \left\{\frac{1-f(v_3)}{1-y}\right\}^{-1} \right\}$$

$$\frac{\partial H}{\partial v_3} = a\left\{ y\left(f(v_3)-1\right) + (1-y)f(v_3) \right\} = a\left\{ f(v_3) - y \right\}$$

$$\frac{\partial H}{\partial w_{12}} = a\left(f(v_3)-y \right) w_{31}\, a f(v_1)\left(1-f(v_1)\right) x_2$$

$$\rightarrow \boxed{\frac{\partial H}{\partial w_{12}} = a^2\left(y^f-y\right) w_{31}\, x_2\, h_1(1-h_1)}$$

مسئله سوم) در یک محیط محاسباتی فقط تابع فعالیت ReLu تعریف شده است؛ اما در لایه آخر (تمام اتصال) نیاز به تابع فعالیت tanh داریم،

الف) پیشنهاد شده است که با استفاده از تابع ReLu تابع tanh را با دقت مناسب ایجاد و استفاده کنیم. این کار چگونه ممکن است؟ **(۱۰)**

تخمین تابع tanh با استفاده از تابع ReLu (از طریق پیش آموزش یک شبکه عصبی)

ب) یک شبکه خیلی ساده با شباهت نسبی به این منظور طراحی کنید (لازم است وزنها و بایاسها و ... را بطور عددی مشخص کنید). منظور یادگیری وزنها نمی باشد، بلکه باید مقادیری برای وزنها و بایاسها پیشنهاد دهید. **(۲۰)**

تابع tanh(x) در 3- حدوداً برابر ۱- و در ۳ حدوداً برابر یک است و مشتق آن در مبدا برابر یک است. بسته به نوع تقریب (حفظ مشتق در مبدا) یا مقدا در کرانهای اشباع می توان آن را با تابع خطی/متقارن/اشباع شونده به شرح زیر تخمین زد. (a=1 or 3)

$$\tanh(x) \simeq \begin{cases} -1 & x < -a \\ \frac{1}{a}x & -a \le x \le a \\ 1 & x > a \end{cases} = -1 + \frac{1}{a}\text{ReLu}(x+a) - \frac{1}{a}\text{ReLu}(x-a)$$

مشخصات شبکه:

یک ورودی

یک لایه مخفی حاوی دو نرون با تابع فعالیت ReLu

یک خروجی با تابع فعالیت خطی