

● به طور کلی (در شکلهای دارای بیش از یک واحد مخفی)، آن وزن های تمام را داده ای مخفی را با صفحه مقادره هی اولیه کنیم، تمام داده ها

مستقل از حعادیر و روای صفر می شوند. بنابراین تمام آن ها دارای تراویان مسیمه به هم فواید ندارند و به این ترتیب تنها اندازه هی بردار وزن ها

متاثر می شود و نه جهت آن به طور کلی مقادره هی اولیه وزن ها به صورت مساوی (صفر)، یادگیری شبکه را از طریق الگوریتم

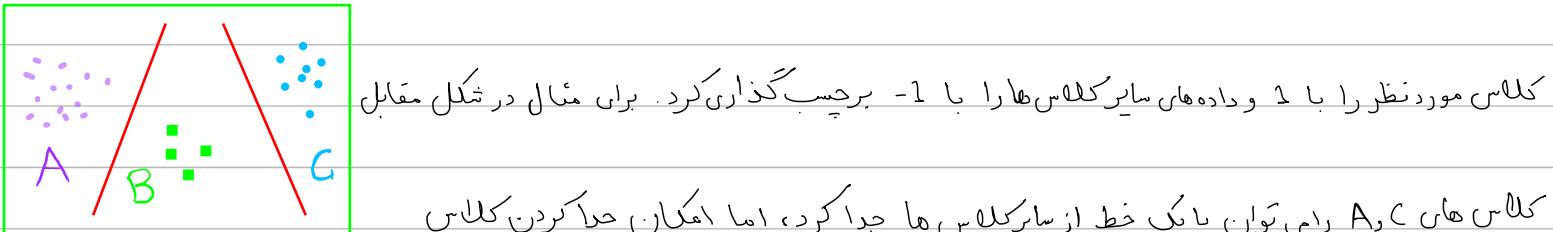
پس انتشار خطا، با مسئله می کند، چرا که خط، متناسب با مقدار وزن ها در شبکه پس انتشار دارد می شود و بنابراین تمام وزن ها به

یک اندازه به روز رسانی می شوند.

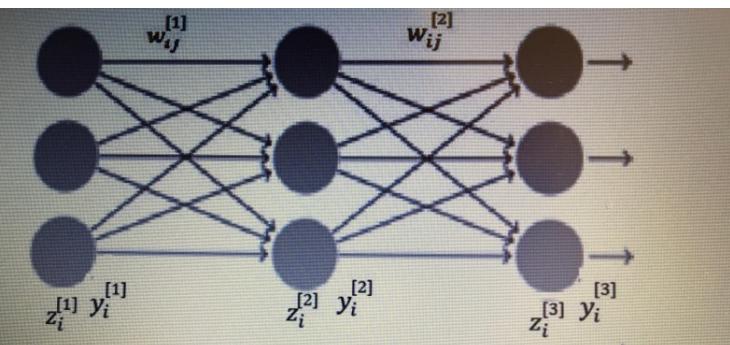
● از random_State برای تدارک زایین ایجاد وزن های تصادفی استفاده می شود. به این ترتیب، در اجراء مختلف الگوی ایجاد وزن ها

تصادفی یکسان خواهد بود و جواب های الگوریتم در هر اجرا (برای داده های تکراری)، یکسان خواهد بود.

می توان از روش All vs 1 استفاده کرد. برای این منظور می توان مدل را به ازای هر کلاس جداگانه آموزش داد و در هر آموزش، داده های



B تنها با یک خط وجود ندارد، لذا می توان داده هایی که جزو هیچ کدام از دیگر کلاس ها نیستند را ب عنوان داده های عضو کلاس B در نظر نرفت.



$$y_i^{[k]} = g(z_i^{[k]})$$

نرون های لایه i قبل از
نرون های لایه $i+1$

$$z_i^{[k]} = \sum_j w_{ij}^{[k-1]} y_j^{[k-1]}$$

عبارت $\frac{\partial E}{\partial z_i^{[k]}}$ ، نکات مختلف نایری \times یکسان است و اساساً تفاوتی در محاسبه دو عبارت $\frac{\partial E}{\partial z_2^{[1]}}$ و $\frac{\partial E}{\partial z_3^{[1]}}$ وجود ندارد.

لایه K ، لایه‌ی آخر باشد

$\frac{\partial y}{\partial z_i} = y'(z_i) = g(z_i)(1-g(z_i))$

در صورتی که مفهوم را زیر می‌بینید در تابع $g(z)$ sigmoid باشد. در سوال صریحاً ذکر نشده است

خطای نرون i نام

آخر خطای نرون i نام را به صورت $e_i = d_i - y_i$ می‌دانیم

آخر خطای لایه i به ازای ورودی n به صورت

$$\frac{\partial E(n)}{\partial z_i^{[2]}} = \underbrace{\frac{\partial E(n)}{\partial e_i(n)}}_{e_i} \cdot \underbrace{\frac{\partial e_i(n)}{\partial y_i(n)}}_{-1} \cdot \underbrace{\frac{\partial y_i(n)}{\partial z_i(n)}}_{\text{ترادیان محل}} = -e_i y'(z_i) = -\delta_i$$

ترادیان محل

$y'(z) = g(z)(1-g(z))$ (با معنی sigmoid)

$$\frac{\partial E(n)}{\partial z_i^{[k<3]}} = -\frac{\partial E(n)}{\partial y_i^{[n]}} \cdot \underbrace{\frac{\partial y_i^{[n]}}{\partial z_i^{[n]}}}_{g'(z)}$$

$$\frac{\partial E(n)}{\partial z_i^{[k<3]}} = \sum_{j \in [k+1]} e_j \frac{\partial e_j(n)}{\partial y_j(n)} = \sum_{j \in [k+1]} \underbrace{\frac{\partial e_j(n)}{\partial z_j(n)}}_{-\delta_j(n)} \cdot \underbrace{\frac{\partial z_j(n)}{\partial y_j(n)}}_{w_{ji}(n)}$$

$$= -\sum_{j \in [k+1]} e_j(n) y'(z_{i(n)}) w_{ji} = -\sum_{j \in [k+1]} \delta_j(n) w_{ji}$$

$$\frac{\partial E(n)}{\partial z_i^{[x<3]}} = -y(z_i) \leq \sum_{j \in [k+1]} \delta_j w_{ji} = -\delta_i$$

لایه i مخفی

$$2) \frac{\partial E}{\partial w_{ij}^{[2]}} = \frac{\partial E}{\partial z_i^{[2]}} \cdot \frac{\partial z_i^{[2]}}{\partial w_{ij}^{[2]}} = -\delta_i^{[2]} y_i^{[2]} = -(y(z_i) \sum_{j \in [3]} \delta_j w_{ji}^{[3]}) y_i^{[2]}$$

منظور مفهوم نرون های لایه $i+1$ از $j=1$ تا 3 است

$$3) \frac{\partial E}{\partial w_{ij}^{[1]}} = -\delta_i^{[1]} y_i^{[1]} = -\left(y'(z_i^{[1]}) \sum_{j \in [2]} \delta_j^{[2]} w_{ji}^{[2]} \right) y_i^{[1]}$$

ب) شبکه های عصبی با هر تعداد لایه را در صورتی که دارای تابع فعالیت خطی باشند، می‌توان به صورت سلسلی

تک لایه بیانه سازی کرد، زیرا در نهایت خروج ها ترکیب خواهند شد و در صورتی که رابطه‌ی بین

ورودی و خروجی‌های آموزشی خطل نباید، سیکل ناکار آمد خواهد بود و اساساً احتفاظ کردن لایه طبق آنچه لفته شد،

بی فایده خواهد بود.

تمرین ۴

$$h_w(x) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3^2$$

تعداد نمونه‌های آموزشی

$$f(w) = \sum_{i=1}^n (y^{(i)} - h_w(x^{(i)}))^2$$

$$f(w) = \frac{1}{2} \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)})^2$$

تایپ هزینه را برای ساده سازی به این \leftarrow سُلّل باز نویسی کرد ایم (سینه کردن این تایپ، باید کمینه شدن تابع فوق می‌شود)

$$\frac{\partial}{\partial w_0} f(w) = \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)})$$

$$\frac{\partial}{\partial w_1} f(w) = \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)}) x_1^{(i)}$$

$$\frac{\partial}{\partial w_2} f(w) = \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)}) x_2^{(i)}$$

$$\frac{\partial}{\partial w_3} f(w) = \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)}) (x_3^{(i)})^2$$

۶ رابط عنوان نزخ یادگیری تعریف می‌کنیم

Repeat {

$$w_j = w_j - \eta \frac{\partial}{\partial w_j} f(w)$$

$j = 0, 1, 2, 3$

مقداری به دست آمده در دست چیز را جایگزنداری می‌کنیم

نکته: تمام پارامترها باید همزمان بروزرسانی شوند.