تمرین 6.2 هوش مصنوعی

همفکران: علیرضا خراسانی

سوال اول) Regression

$$\min_w f = \lambda w^T w + \| Xw - Y \|^2$$

آ) از overfit شدن method به داده های train جلوگیری می کند.

در واقع با اضافه کردن $\lambda w^T w$ ضرایب w_i کوچک خواهند شد و از تغییرات زیاد تابع جلوگیری می کند.

ب)

$$\min f(w) = \lambda w^T w + \| Xw - Y \|^2 = \lambda w^T w + (Xw - Y)^T (Xw - Y)$$

$$\frac{\partial F}{\partial w} = \cancel{2} \lambda w + \cancel{2} X^T (Xw - Y) = 0 \Rightarrow$$

$$(\lambda + X^T X) w = X^T Y \Rightarrow$$

$$\boxed{w = (\lambda I + X^T X)^{-1} X^T Y}$$

دقت کنید که X یک ماتریس از داده هاست.

ج)

$$\begin{cases} w_1 = \arg\min_w L(w) \\ w_2 = \arg\min L(w) + \lambda w^T w \end{cases} \Rightarrow L(w_1) \leq L(w_2)$$

$$w_2 = \arg\min L(w) + \lambda w^T w \Rightarrow L(w_2) + \lambda w_2^T w_2 \leq L(w_1) + \lambda w_1^T w_1$$

$$\Longrightarrow \cancel{L(w_1) + L(w_2)} + \lambda w_2^T w_2 \leq L(w_1) + \cancel{L(w_2)} + \lambda w_1^T w_1$$

$$\Longrightarrow \|w_2\| \leq \|w_1\|$$

← همانطور که مشخص است.

w_2 از w_1 اندازه کمتری دارد ⟸ w_i ها کوچکتر اند [در حالت کلی].

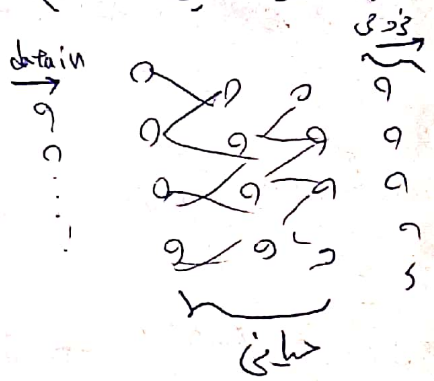
Neural Network

1

می‌دانیم که خروجی هر نورون به صورت $a w^T x + b$ و ورودی نورون‌های لایه‌های میانی و لایه آخر

خروجی نورون‌های قبل است ⟸

ورودی هر نورون (بجز نورون لایه اول) $= W_{n-1}^T X_{n-1}$



خروجی نورون $= a_n W_{n-1}^T (a_{n-1} W_{n-2}^T X_{n-2}) + b_n$

$X_n = a_n W_{n-1}^T X_{n-1} + b_n$

خروجی نورون n ام.

پس خروجی هر نورون را می‌توان به صورت ترکیب خطی از خروجی نورون‌های

لایه قبل نوشت ⟸ بنابراین می‌توان لایه‌های میانی را حذف کرد.

$W_n = [w_{n_1} \dots b_n]$

$X_n = [x_{n_1} \dots 1]$

دقت کنید که می‌توان بایاس هر لایه و وزن را به وسیله a و b تاثیر داد؟

$X_n = W_{n-1}^T X_{n-1} \quad +$

بنابراین می‌توان به روش بازگشتی لایه‌های میانی را حذف کرد و یک لایه ورودی و خروجی

$X_0 = data$

بدست آوریم.

2

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial y} \times \frac{\partial y}{\partial b}$$

$$y = f(wx + b) \rightarrow \frac{\partial y}{\partial b} = \frac{\partial f(wx+b)}{\partial (wx+b)} \times \overset{1}{\frac{\partial (wx+b)}{\partial b}} = \frac{\partial f(x)}{\partial x} =$$

$$\rightarrow \frac{\partial E}{\partial b} = \frac{\partial E}{\partial y} \times f'$$

$$\frac{\partial E}{\partial W}$$

محاسبه

دقت کنید که اگر X ماتریس باشد. آنگاه

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{m1}} & & \frac{\partial y}{\partial x_{mn}} \end{bmatrix}$$

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial y} \times \frac{\partial y}{\partial W}$$

$$\Rightarrow \frac{\partial E}{\partial W} = \frac{\partial E}{\partial y} \times \frac{\partial f(wx+b)}{\partial f(wx+b)} \times \frac{\partial (wx+b)}{\partial W} = \frac{\partial E}{\partial W} = \frac{\partial E}{\partial y} \times f' \times X$$

$\Rightarrow$ جمع بندی اگر $u = wx+b$ بگیریم.

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial y} \cdot \frac{\partial f}{\partial u} \cdot 1$$

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial y} \cdot \frac{\partial f}{\partial u} \cdot X$$