Date Subject

HW7

یاسین عسگریان

بطور کلی overfit شدن یعنی حفظ کردن داده های آموزشی

و underfit شدن یعنی یک مدل ساده تر اما الگوهای کمتری

از داده های آموزشی استخراج کرده و دقت هم روی داده های آموزشی

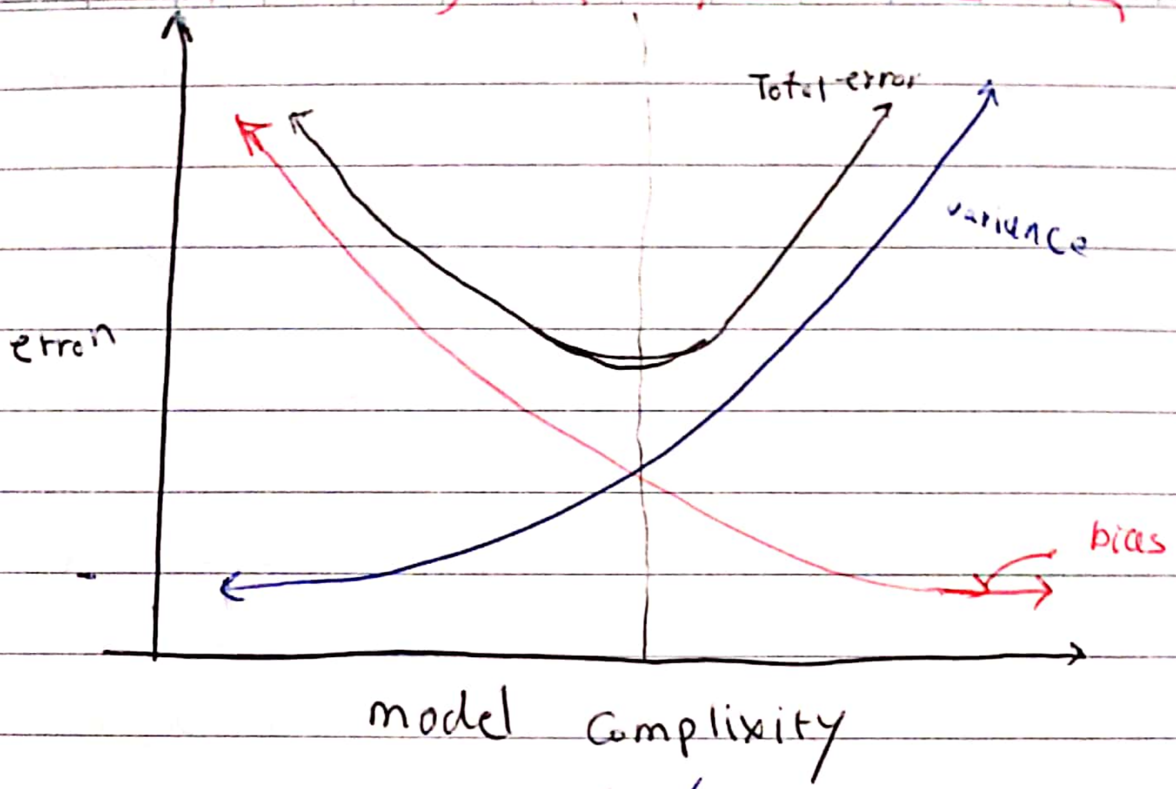
هم روی دیتای تست پایین است

ما به دنبال این هستیم که low bias و low variance داشته باشیم

برای بیان ارتباط بین bias و variance با overfitting

و underfitting به موارد زیر توجه فرمایید

Date undefitting Subject overfitting



همانطور که در نمودار مشاهده میکنید اگر variance زیاد باشد و bias

کم باشد مدل overfit میشود چون خروجی مدل بطور میانگین

بسیار نزدیک به هدف است اما پراکندگی خروجی ها زیاد است یعنی

با تغییر دستی از فیچرها خروجی متفاوتی دریافت میکنیم

برای underfitting ما bias بالا و واریانس کم داریم یعنی

خروجی مدل به صورت میانگین از هدف دور است و پراکندگی خروجی ها

CHITRA

برروی آن میانگین خطا بیشتر است

~~برای overfitting~~

برای جلوگیری از overfit شدن می‌توانیم از مدل ساده‌تر استفاده کنیم (نه آنقدر که underfit شود) از ~~[illegible]~~ منظم‌سازی های L1 یا L2 استفاده کنیم با استفاده از Dropout بعضی از نورون ها را نادیده بگیریم در ترین

برای جلوگیری از underfit شدن مدل را باید مقداری پیچیده‌تر کنیم ~~[illegible]~~ و اگر دیتاهای ما کم است یا نامتوازن بهتر است از تکنیک های داده افزایی برای افزایش داده هایمان استفاده کنیم (~~[illegible]~~ داده افزایی همه جا مناسب است)

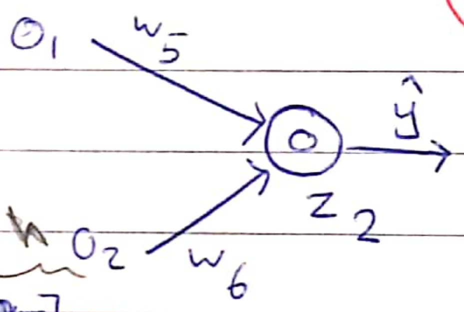
در کل بهتر است از این روش ها متناسب با مسئله مان بهره‌گیری

استفاده کنیم. با استفاده از روش های مثل k-fold به

اعتبارسنجی Hyperparametr ~~Hyper~~ ها و مدل های متفاوت

بپردازیم تا بهترین مدل را استخراج کنیم

ج ۲)



$$Z_2 = \overbrace{[w_5, w_6]}^{w^{[2]}} \overbrace{\begin{bmatrix} O_1 \\ O_2 \end{bmatrix}}^{h} + b_2$$

$$E = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\hat{y} = \text{Sigmoid}(z_2)$$

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_2} \cancelto{h}{\frac{\partial z_2}{\partial w}}$$

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_2} \cancelto{1}{\frac{\partial z_2}{\partial b}}$$

Date Subject

$$\frac{\partial E}{\partial \hat{y}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$= \frac{\partial(y^2 + \hat{y}^2 - 2y\hat{y})}{\partial \hat{y}} = 2\hat{y} - 2y = -2(y-\hat{y})$$

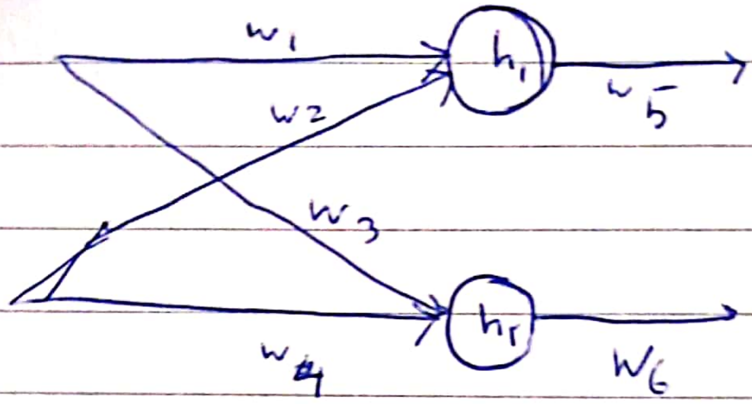
$$= \frac{\partial E}{\partial y} = -2(y-\hat{y}) \Rightarrow \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

$$\frac{\partial \hat{y}}{\partial z_3} = \sigma(z_3)(1-\sigma(z_3))$$

$$\frac{\partial E}{\partial w} \Rightarrow \frac{-2}{m}\sum_{i=1}^{n}(y_i - \hat{y}_i)\,\sigma_i(z_2)(1-\sigma_i(z_2))h_i$$

$$\frac{\partial E}{\partial b} = \frac{-2}{m}\sum_{i=1}^{n}(y_i - \hat{y}_i)\,\sigma_i(z_2)(1-\sigma(z_2))$$

Date Subject



$$Z_1 = I \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \end{bmatrix}^T \underbrace{\begin{bmatrix} i_1 & i_2 \\ i_1 & i_2 \end{bmatrix}}_{I} + \begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix}$$

$$E = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\hat{y} = \text{Relu}(Z) \quad \begin{cases} 1 & Z > 0 \\ 0 & Z < 0 \end{cases}$$

$$\frac{\partial E}{\partial w^{[1]}} = \frac{\partial E}{\partial \hat{y}} \cancel{\frac{\partial \hat{y}}{\partial h}} \cancel{\frac{\partial h}{\partial z_1}} \cancel{\frac{\partial z_1}{\partial w^{[1]}}} \quad I$$

$$\cancel{\frac{\partial \hat{y}}{\partial z}} \; \cancel{\frac{\partial z}{\partial h}} \rightarrow w^{[2]}$$

$$\hat{y}(1-\hat{y})$$

$$\frac{\partial E}{\partial w^{[1]}} \begin{cases} -\frac{2}{m}(y-\hat{y})\,\hat{y}(1-\hat{y})\,w^{[2]}\, I & Z_1 > 0 \\ 0 & Z_1 < 0 \end{cases}$$

CHITRA

$w_1 = 0{,}5$, $w_2 = -0{,}5$, $w_3 = -3$, $w_4 = 1{,}5$

$w_5 = -1{,}5$, $w_6 = 1$

$b_1, b_2, b_3 = 0$

Itraition 1

$8 \longrightarrow 0$

$20 \longrightarrow 1$

Layer1

$i_1 = 3$

$i_2 = 2$

$$3w_1 + 2w_2 + 0 = 3(,5) + 2(-,5) + 0 = \boxed{0{,}5}$$

$$3w_3 + 2w_4 + 0 = 3(-3) + 2(1{,}5) + 0 = -6$$

$O_1 = \mathrm{Relu}(0{,}5) = 0{,}5$

$O_2 = \mathrm{Relu}(-6) = 0$

Layer2

$$O_1 w_5 + O_2 w_6 + b_3$$

$$\Rightarrow 0{,}5(-1{,}5) + 0 + 0 = -0{,}75$$

$$\hat{y} = \mathrm{Sigmoid}(-0{,}75) \simeq 0.32$$

Date Subject

layer 1

$i_1 = 15$

$i_2 = 12$

$$\begin{cases} i_1 w_1 + i_2 w_2 + b_2 \\ i_1 w_3 + i_2 w_2 + b_2 \end{cases}$$

$$= \begin{cases} 15(0{,}5) + 12(-0{,}5) + 0 = 1{,}5 \\ 15(-3) + 12(1{,}5) + 0 = -27 \end{cases}$$

$$\Rightarrow \begin{cases} Relu(1{,}5) = 1{,}5 \\ Relu(-27) = 0 \end{cases}$$

layer 2

$$O_1 w_5 + O_2 w_6 + b_3$$

$$\Rightarrow 1{,}5(-1{,}5) + 0 + 0 = -2.25$$

$$\hat{y} = Sigmoid(-2.25) \simeq 0.09$$

back propagation

$w_5 = -0.014$

$w_6 = 0$

$$d_w = -\frac{٢}{٢}\left(\left((0 - 0.32)\, \cancel{0.2}\, 0.32(1 - 0.32)\begin{bmatrix} 0.5 \\ 0 \end{bmatrix}\right) + \right.$$

ناستام

$$\left.\left((1 - 0.9)(0.9(1 - 0.9)\begin{bmatrix} 0.5 \\ 0 \end{bmatrix}\right)\right)$$

$w_5 = \cancel{0.023}\ 0.059$

$w_6 = 0$

$dw_5 = -0.045$

$dw_6 = 0$

$$\boxed{db_3 \simeq 0.005}$$

Apply L_2 norm

$$\tilde{J}(\theta; x, y) = \frac{\alpha}{2} w^T w + J(\theta; x, y)$$

$$\nabla_w \tilde{J}(\theta; x, y) = \alpha w + \nabla_w J(\theta; x, y)$$

$\alpha = 0.01$

$dw_5 = (0.01)(-1.5) + 0.045 = -0.06$

$dw_6 = (0.01)\,1 + 0 = 0.01$

اپدیت کردن Adam بصورت که انجام شد

$w_5 = -1.49$

$b_3 = 0.0005$

$w_6 = 1$

$$dw_1 = \frac{-2}{2}\Big[\big((0-0.32)(0.32)(1-0.32)\,\overset{-1.5}{\cancel{w_5}}\;\overset{3}{\cancel{i_1}}\big) + \big((1-0.9)(0.9)(1-0.9)\,w_5\;\overset{15}{\cancel{i_1}}\big)\Big] = +1.36$$

$$dw_2 = \frac{-2}{2}\Big[\big((0-0.32)(0.32)(1-0.32)\,w_5\;\overset{2}{\cancel{i_2}}\big) + \big((1-0.09)(0.09)(1-0.09)\,w_5\;\overset{12}{\cancel{i_2}}\big) = 1.27$$

$$dw_3 = \frac{-2}{2}\Big[\big((0-0.32)(0.32(1-0.32))\,\overset{1}{\cancel{w_6}}\;\overset{3}{\cancel{i_1}}\big) + \big((1-0.9)(0.9(1-0.9))\,w_6\;\overset{15}{\cancel{i_1}}\big)\Big] = -0.9$$

$$dw_4 = \frac{-2}{2}\Big[(0-0.32)(0.32(1-0.032)\,w_6\;\overset{2}{\cancel{i_2}} + \big((1-0.09)(0.09(1-0.09)\,w_6\;\overset{12}{\cancel{i_2}}\big)\Big]$$

$$= 0.742$$

بر db همان فرمول dw اعمال می‌شود مثلاً

I ماندارد

اعمال L2 نرمال

$dw_1 = (0.01)(0.5) + 1.36 = 1.365$

$dw_2 = (0.01)(-0.5) + 1.27 = 1.265$

$dw_3 = (0.01)(-3) - 0.9 = -0.93$

$dw_4 = (0.01)(1.5) + 0.42 = 0.435$

بعد از اعمال Adam داریم

$w_1 = 0.49$ $\qquad db_1 = 0.21$

$w_2 = -0.50099$ $\qquad db_2 = 0.005$

$w_3 = -3.0009$ $\qquad b_1 = -0.0009$

$w_4 = 1.49$ $\qquad b_2 = -0.0009$

Date Subject

* برای تکرار دهم هم مقادیر نزدیک به همین مقادیر است و به نظر می‌رسد برای همگرا شدن به تکرارهای بیشتری نیاز شده است

ج ۳)

الف)

ابتدا ۴ نوع مدل با آموزش Tiny، Small، medium و large استفاده کرد که اسم های مدل از پیچیدگی مدل گرفته شده است که مشاهده شد به غیر از Tiny در بقیه overfitting رخ داده است ~~large~~

ابتدا روی مدل large از L2 norm استفاده کرد که دقت خیلی آمد پایین اما مدل دیگر overfit نیست

لذا dropout استفاده کرده و در نهایت هم از dropout

هم از L2norm استفاده کرده است که هم overfit و هم underfit

نشده است هم به دقت 72 درصد رسیده است

البته روی این مدل نتوانست دقت را زیاد افزایش دهد

مثل large روی داده‌ها Train به دقت 100 رسید

جلو رفته بود که البته overfit شده بود اما به طور کلی .

Regularization هایی که انجام داد که به جلوگیری از overfit

شدن کمک کرد و با L2 + ~~[illegible]~~ Dropout از overfit شدن

جلوگیری کرد

ب)

ابتدا سعی کردم تعداد نورون‌ها های هر لایه را یکی بگذارم

و نورون‌ها در هر لایه افزایش یابند در هم تلاش کردم Dropout

CHITRA

Date Subject

برای همه لایه‌ها بار نرم اما متوجه شدم با پیچیده‌تر شدن

مدل کماکان مدل به سمت overfitting حرکت می‌کند و

نیاز به منظم‌سازی می‌باشد در نهایت PReLU را اضافه کردم

بدون Normalization هم مدل overfit شد و

و PRelu Normalize کردم

در نهایت مدل آخر Custom4 به

Validation Acc = 68%

Acc = 71%

رسید

به تصویر مربوط به این سوال پیوست شده است

که مربوط به سوال ۲ پیوست شده است