①

الف)

راهکار ۱- استفاده از مدل پیچیده‌تر- به این منظور از underfit شدن که موجب ارور زیاد است بیشتر جلوگیری می‌کنیم.

راهکار 2- (بهبود) استفاده از feature Engineering یا افزایش داده (در بعضی مواقع)، دقیق‌تر یاد گرفته که موجب بدین وسیله پترن های مهم تر داده را با میزان داده کمتر کاهش بایاس مدل و جلوگیری از underfitting هم می‌شود.

ب) وابستگی داشتن موجب می‌شود تأثیری که این متغیرها دارند بیشتر محسوب بگردد. که هم موجب بایاس می‌شه چون تأثیر سایر متغیرها دارد کم در نظر گرفته می‌شود همچنین با تغییر کمی در متغیرهای همبسته تأثیر زیادی در نتیجه مدل رخ می‌دهد که موجب واریانس بالا است.

در صورت حذف کلی این متغیرها ممکن است information loss رخ دهد که به موجب آن bias افزایش یابد (و مدل به خوبی شناخته نشده) و می‌تواند موجب کاهش واریانس شود چرا که تنوع ورودی‌ها را کاسته ایم.

ج) اولی درسته

اولی: اگر تعداد مدل کم باشد که بایاس از منشأ underfit بوده باشد، افزایش داده‌ها تأثیری ندارد. ولی می‌تواند در بعضی حالات که ناشی از کمبود دیتا بوده باشد، تأثیر مثبت داشته باشد.

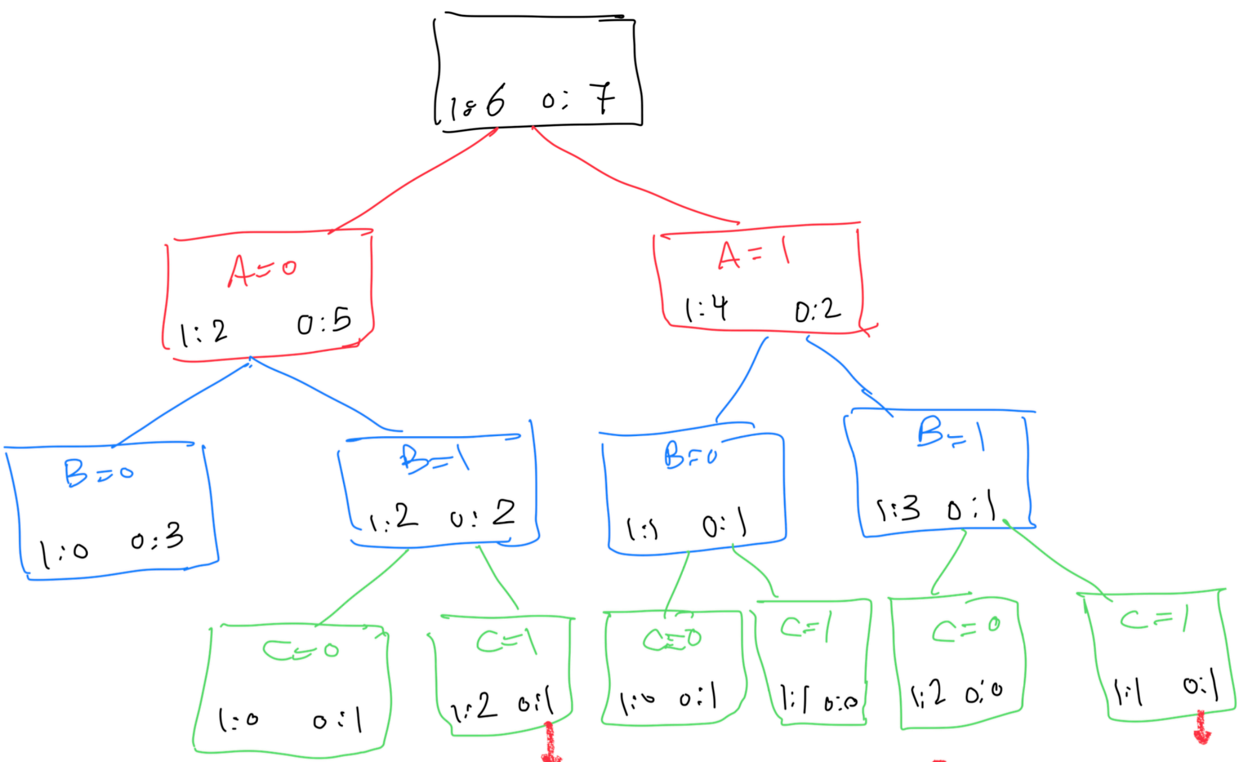
دومی: همواره نه. در حالاتی که منشأ overfit نیست و منشأ underfit است با افزایش پیچیدگی می‌تواند موجب کاهش هر دو خطا بشود.

② الف) داده‌هایی که تمام فیچرها مقدار برابر دارند، در صورت اختلاف خروجی، موجب خطایمان اند.
آنهایی که با عمومیت خروجی دسته فیچر برابرهاشان متناقض اند، غلط پیش‌بینی می‌شوند.

C	B	A	X	Pred	isWrong
0	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
0	1	0	0	0	0
1	1	0	0	1	1
1	1	0	1	1	0
1	1	0	1	1	0
0	0	1	0	0	0
1	0	1	1	1	0
0	1	1	1	1	0
0	1	1	1	1	0
1	1	1	0	0	0
1	1	1	1	0	1

رندوم بین ۰ را سه داریم

خطا $\frac{2}{13} = 15.4\%$



خطا: 2
کل: 13

پرهام رضایی ۴۰۰۱۰۸۵۴۷

ب)

ماکسیمم خطا زمانی است که هیچ فیچر جدا کننده ای نداریم.

n تا داده و K تا کلاس

در زمانی که فیچر همه داده ها یکیه حداقل $\lceil \frac{n}{K} \rceil$ تا درست پیش بینی میشه

و خطا میشه $\frac{\text{نست}}{\text{کل}} = \frac{n - \lceil \frac{n}{K} \rceil}{n}$

با استقرا نشان میدم ✱ درسته.

پایه برای $n=2$: اگر فیچر جدا کننده داشته باشیم پس ۱۰۰٪ جدا میشوند ولی

در غیر این $50\% \leftarrow$ ✓

گام استقرا: برای $m<n$ درسته برای n هم درسته.

که بدون جدا کننده نمیشد. مشابه (a)

فرض خلف میکنیم که فیچر جدا کننده داریم و حداکثر خطا رسیدیم.

فرض میکنیم در دسته i ام P_i تا داده اند پس درست ها حداقل $\lceil \frac{P_i}{c} \rceil$ ام آن

اند و در کل $\sum_i \lceil \frac{P_i}{c} \rceil$ ها درست داریم.

اگر فیچر جدا کننده نداشتیم ولی $\lceil \frac{n}{c} \rceil$ تا حداقل درست بود

پس باید طبق فرض خلف $\sum_i \lceil \frac{p_i}{c} \rceil < \lceil \frac{n}{c} \rceil$

$$\Rightarrow \quad \sum_i \left\lceil \frac{p_i}{c} \right\rceil < \left\lceil \frac{\sum p_i}{c} \right\rceil$$

که تناقض است چون

$$\sum_i \left\lceil \frac{x_i}{c} \right\rceil \geq \left\lceil \frac{\sum x_i}{c} \right\rceil$$

اثبات: $x_i = r_i c + \alpha_i$

$0 \leq \alpha_i < c$

$$\Rightarrow \quad LHS = \sum_i r_i + \sum_i I(\alpha_i > 0)$$

$$RHS = \sum_i r_i + \left\lceil \frac{\sum \alpha_i}{c} \right\rceil$$

ادعا:

$$\sum_i I(\alpha_i > 0) \geq \left\lceil \frac{\sum \alpha_i}{c} \right\rceil$$

بدیهی است چرا که اگر همه ۰ که برابرند

اگر s تا $\alpha_i > 0$ باشند:

$$LHS = s \qquad RHS = \left\lceil \frac{\sum^{s} \alpha_{i_j}}{c} \right\rceil$$

که

$$\sum_{j=1}^{s} \alpha_{i_j} < sc$$

$$\Rightarrow \quad RHS \leq \left\lceil \frac{sc}{c} \right\rceil = s \quad \blacksquare$$

در حالت کلی فارغ از تعداد دیتا (یعنی بدون حسب تعداد دیتا ماکسی که میتونه رخ بده) به شرح زیر: {جواب اصلی همون بالاییه}

ب)

یک دسته را در نظر می‌گیریم که همگی مقادیر فیچرهای برابر دارند. خروجی مدل برابر خروجی اکثریت آنها است. پس در حالت غیرمتساوی اگر t تا ماکسیمم کلاس باشند، $t-1$ تا از سایر کلاس‌ها $(t-1)(k-1)$ تا عضو خطا داریم که می‌شود $\frac{(t-1)(k-1)}{(t-1)(k-1)+t} = \frac{tk-t-k+1}{tk-k+1}$ که کامل نیست چون حالت تساوی هم داریم. اگر همه t باشند و یکی را انتخاب کنیم $\Rightarrow$ $t(k-1)$ تا خطا اند.

$$\frac{t(k-1)}{t(k-1)+t} = \frac{k-1}{k}$$

حال جمع این‌ها برای تمام دسته‌ها چون نسبت به طول ثابت است برابر $\frac{k-1}{k}$ کل است.

$$\frac{\sum_{\ell} \frac{k-1}{k}}{|\ell|} = \frac{\frac{(k-1)|\ell|}{k}}{|\ell|} = \boxed{\frac{k-1}{k}}$$ جواب نهایی

3

الف)

$$P(Y=k \mid X=\langle X_1,\dots,X_n\rangle) = \frac{\exp(w_k^T X + b)}{\sum_j \exp(w_j^T X + b)}$$

$$P(Y=K \mid X=x) + \sum_{k=1}^{K-1} P(Y=k \mid X=x) = 1$$

$$\Rightarrow P(Y=K \mid X=x) = 1 - \sum_{k=1}^{K-1} P(Y=k \mid X=x) \overset{(*)}{=} 1 - \sum_{k=1}^{K-1} P(Y=K \mid X=x)\, e^{w_k^T x}$$

$$\Rightarrow P(Y=K \mid X=x) = \frac{1}{1+\sum_{k=1}^{K-1} e^{w_k^T x}} \Rightarrow P(Y=k \mid X=x) = \frac{e^{w_k^T x}}{1+\sum_{k=1}^{K-1} e^{w_k^T x}}$$

$(*)$: $P(Y=i \mid X=x) = P(Y=K \mid X=x)\, e^{w_i^T x}$ $\quad i<K$

$$\ln \frac{P(Y=i \mid X=x)}{P(Y=K \mid X=x)} = e^{w_i^T x} \rightarrow$$

چرا که این را خاص گذاشتیم additive log ratio transform

$K-1$ تا باید لامبدا بزنیم یکی رو خالی گرفتیم (K امی)

ب) $w_1, \dots, w_{K-1}$ باید تخمین زده شوند. (پارامترها) ← هر یک m بعدی باشند ← $(K-1)m$

K های بی پارامتره و در این رگرسیون ثابته.

ب)

$$L(w_1,\dots,w_{K-1}) = \sum_{i=1}^{n} \ln P(Y=y_i \mid X=x_i) =$$

$$\sum_{i=1}^{n} \ln \frac{e^{w_{y_i}^T x_i}}{1+\sum_{j=1}^{K-1} e^{w_j^T x_i}} = \sum_{i=1}^{n} \ln(e^{w_{y_i}^T x_i}) - \sum_{i=1}^{n} \ln\left(1+\sum_{j=1}^{K-1} e^{w_j^T x_i}\right)$$

در تابع‌ها یکی هم میشه کرد

$$= \sum_{i=1}^{n} w_{y_i}^T x_i - \sum_{i=1}^{n} \ln\left(1+\sum_{j=1}^{K-1} e^{w_j^T x_i}\right)$$

ج)

$$(\blacksquare):\quad \frac{\partial L(w_1,\dots,w_{K-1})}{\partial w_t} = \sum_{i=1}^{n} \mathbb{I}(y_i=t)\, x_i - \sum_{i=1}^{n} \frac{x_i e^{w_t^T x_i}}{1+\sum_{j=1}^{K-1} e^{w_j^T x_i}} \quad \circledast$$

$$\circledast:\quad \frac{\partial}{\partial w_t} \sum_{j=1}^{K-1} e^{w_j^T x_i} = 0 + \frac{\partial}{\partial w_t} e^{w_t^T x_i} = e^{w_t^T x_i} \frac{\partial w_t^T x_i}{\partial w_t} = x_i e^{w_t^T x_i}$$

د)

(هرکدوم اوکی) بالاییست

$$\frac{\partial f(w_1,\dots,w_{K-1})}{\partial w_t} = (\blacksquare) - \frac{\lambda}{2}\frac{\partial}{\partial w_t}\sum_{j=1}^{K-1} \|w_j\|_2^2 = (\blacksquare) - 0 - \frac{\lambda}{2}\frac{\partial\, w_t^T w_t}{\partial w_t} =$$

$$(\blacksquare) - \frac{\lambda}{2}(2w_t) = (\blacksquare) - \lambda w_t$$

پرهام رضایی ۴۰۰۱۰۸۵۴۷

ب) با تعریف $I_{ik} \leftarrow \begin{cases} 1 & y_i = k \\ 0 & o.w \end{cases}$ داریم:

$$L(w_1, \ldots, w_{k-1}) = \sum_{i=1}^{n} \sum_{j=1}^{k} I_{ij} \ln P(Y=j \mid X = x_i)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k} I_{ij} \left(w_j^T x_i - \ln \left(\sum_{j=1}^{k} e^{w_j^T x_i} \right) \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} I_{ij} w_j^T x_i - \sum_{i=1}^{n} \sum_{j=1}^{k} I_{ij} \ln \left(\sum_{j=1}^{k} e^{w_j^T x_i} \right))$$

(ت

$$\frac{\partial}{\partial w_t} \sum_{j=1}^{n} w_j^T \sum_{i=1}^{n} I_{ij} x_i = 0 + \frac{\partial}{\partial w_t} w_t^T \sum_{i=1}^{n} I_{it} x_i = \sum_{i=1}^{n} I_{it} x_i$$

$$\frac{\partial}{\partial w_t} L(w_1, \ldots, w_{k-1}) = \sum_{i=1}^{n} I_{it} x_i - \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{I_{ij} x_i e^{w_t^T x_i}}{\sum_{j=1}^{k} e^{w_j^T x_i}}$$

$$= \sum_{i=1}^{n} \left(I_{it} - \sum_{j=1}^{k} I_{ij} P(Y=j \mid X = x_i) \right) x_i =$$

$$\sum_{i=1}^{n} \left(I_{it} - P(Y = y_i \mid X = x_i) \right) x_i$$

④

الف)

مشابه رگرسیون عادی ولی ورودی فقط فیچر j ام است

$$\Rightarrow L = \frac{1}{2}\sum_{i=1}^{n}\left(\underset{\text{عدد ثابت } c}{w^T}\,\underset{1\times 1}{x_{j,i}} - y_i\right)^2$$

$$\Rightarrow \frac{\partial L}{\partial c} = \frac{1}{2}\sum_{i=1}^{n} 2(c x_{j,i} - y_i)\, x_{j,i} = \sum_{i=1}^{n}(c x_{j,i} - y_i)\, x_{j,i} =$$

$$= X_j^T(cX_j - Y)$$

$\Rightarrow$ می‌توان نوشت: $\quad cX_j^TX_j - X_j^TY = 0 \Rightarrow \overset{w_j}{\overset{\shortparallel}{c}} = \dfrac{X_j^TY}{X_j^TX_j}$

ثابت پس جابجایی

عدده پس میشه آورد $(X_j \neq 0 \text{ if})$

ب)

$$\frac{\partial L}{\partial w_j} = \frac{\partial}{\partial w_j}\frac{1}{2}\sum_{i=1}^{n}(w^TX_i - y_i)^2 =$$

$$\sum_{i=1}^{n}(w^TX_i - y_i)\frac{\partial}{\partial w_j}w^TX_i = \sum_{i=1}^{n}(w^TX_i - y_i)\, x_{i,j}$$

صفر شد

$$= \langle w^TX - Y, X_j\rangle = 0 \Rightarrow$$

$$X_j^Tw^TX - X_j^TY = 0 \Rightarrow X_j^Tw^TX = X_j^TY$$

$$\Rightarrow [w_1X_j \cdots w_mX_j]\, X = X_j^TY$$

$$[w_1X_j^T \cdots w_mX_j^T]\begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} = w_1X_j^TX_1 + \cdots + w_mX_j^TX_m = w_j X_j^TX_j$$

$$\Rightarrow w_j X_j^TX_j = X_j^TY$$

$$\Rightarrow w_j = \frac{X_j^TY}{X_j^TX_j}$$

$\underset{i\neq j}{X_j^TX_i = 0}$ طبق تعامد 2

پ)

$$L = \frac{1}{2} \sum_{i=1}^{n} \left(w_j x_{j,i} + w_0 - y_i \right)^2$$

$$\frac{\partial L}{\partial w_j} = \frac{1}{2} \sum_{i=1}^{n} 2 \left(w_j x_{j,i} + w_0 - y_i \right) x_{j,i} = \sum_{i=1}^{n} \left(w_j x_{j,i}^2 + w_0 x_{j,i} - y_i x_{j,i} \right)$$

$$\frac{\partial L}{\partial w_0} = \frac{1}{2} \sum_{i=1}^{n} 2 \left(w_j x_{i,j} + w_0 - y_i \right) = \sum_{i=1}^{n} \left(w_j x_{i,j} + w_0 - y_i \right)$$

$$(0)\ \frac{\partial L}{\partial w_j} = 0 \rightarrow \quad \langle Y, X_j \rangle = \sum_{i=1}^{n} \left(w_j x_{j,i} + w_0 \right) x_{j,i} = \langle w_j X_j + w_0 , X_j \rangle$$

$$\frac{\partial L}{\partial w_0} = 0 \Rightarrow \quad \sum_{i=1}^{n} w_j x_{j,i} + n w_0 = \sum_{i=1}^{n} y_i \quad \xRightarrow{\div n}$$

$$\frac{\sum_{i=1}^{n} w_j x_{j,i}}{n} + w_0 = \frac{\sum_{i=1}^{n} y_i}{n} \Rightarrow \quad \mathbb{E}(w_j x_j) + w_0 = \mathbb{E}(y)$$

$$\Rightarrow \quad w_0 = \mathbb{E}(y) - \mathbb{E}(\underset{\text{ثابت}}{w_j} x_j) = \mathbb{E}(y) - w_j \mathbb{E}(x_j) \quad (\star)$$

$$(0) \rightarrow \quad w_j \sum_{i=1}^{n} x_{j,i}^2 = \sum_{i=1}^{n} y_i x_{j,i} - w_0 \sum_{i=1}^{n} x_{j,i} \quad \xrightarrow{\div n}$$

$$w_j \frac{\sum_{i=1}^{n} x_{j,i}^2}{n} = \frac{\sum_{i=1}^{n} y_i x_{j,i}}{n} - w_0 \frac{\sum_{i=1}^{n} x_{j,i}}{n} \Rightarrow$$

$$w_j \mathbb{E}(x_j^2) = \mathbb{E}(y x_j) - w_0 \mathbb{E}(x_j)$$

$$(\star) \xrightarrow{\text{جای}} w_j \mathbb{E}(x_j^2) = \mathbb{E}(y x_j) - w_0 \mathbb{E}(x_j) = \mathbb{E}(y x_j) - \left(\mathbb{E}(y) - w_j \mathbb{E}(x_j) \right) \mathbb{E}(x_j)$$

$$\Rightarrow \quad w_j \big(\underbrace{\mathbb{E}(x_j^2) - \mathbb{E}(x_j)^2}_{var(x_j)} \big) = \mathbb{E}(y x_j) - \mathbb{E}(y) \mathbb{E}(x_j) = cov[x_j, y]$$

$$\Rightarrow \quad w_j = \frac{cov[x_j, y]}{Var[x_j]}$$

۴) با ناهمبستگی کامل:

الف) معادله دستامون فقط x_j است پس

$$w_j = (x_j^T x_j)^{-1} x_j y = \frac{x_j^T y}{x_j^T x_j}$$

ب) متعامدند ← $x_i^T x_j = 0 \quad i \neq j$

$$\Rightarrow \quad X^T X = \begin{bmatrix} x_1^T x_1 & & 0 \\ & \ddots & \\ 0 & & x_m^T x_m \end{bmatrix}$$

قطری ← معکوس هم قطری

$$\Rightarrow (X^T X)^{-1} = \begin{bmatrix} (x_1^T x_1)^{-1} & & 0 \\ & \ddots & \\ 0 & & (x_m^T x_m)^{-1} \end{bmatrix}$$

$$\Rightarrow w = (X^T X)^{-1} X^T y = \begin{bmatrix} (x_1^T x_1)^{-1} & & 0 \\ & \ddots & \\ 0 & & (x_m^T x_m)^{-1} \end{bmatrix} X^T y$$

$$\rightarrow w_i = (x_i^T x_i)^{-1} (X^T y)_i = \frac{(X^T y)_i}{x_i^T x_i} = \frac{X^T_{i:} y}{x_i^T x_i} = \frac{x_i^T y}{x_i^T x_i}$$

⇐ مستقلاً برای جواب بهینه دید هر متغیر تعیین می‌شود.

$S(X) =$ جمع کل بردار $= x_1 + \dots + x_m$

ضرب دو ماتریس 2×2 درون

ب)

$$X^TX = \begin{bmatrix} x_j^T x_j & S(x_j) \\ S(x_j) & n \end{bmatrix} \Rightarrow (X^TX)^{-1} = \frac{1}{n\|x_j\|^2 - S(x_j)^2} \begin{bmatrix} n & -S(x_j) \\ -S(x_j) & x_j^T x_j \end{bmatrix}$$

det

$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ چون در داخل ضرب تکه

$$w = (X^TX)^{-1}X^Ty = \frac{1}{n\|x_j\|^2 - S(x_j)^2} \begin{bmatrix} n & -S(x_j) \\ -S(x_j) & x_j^T x_j \end{bmatrix} \begin{bmatrix} x_j^T y \\ S(y) \end{bmatrix}$$

مولفه اول

$$\Rightarrow w_j = \frac{n x_j^T y - S(x_j)S(y)}{n\|x_j\|^2 - S(x_j)^2} = \frac{\frac{x_j^Ty}{n} - \frac{S(x_j)}{n}\frac{S(y)}{n}}{\frac{\|x_j\|^2}{n} - \left(\frac{S(x_j)}{n}\right)^2} = \frac{E(x_j y) - E(x_j)E(y)}{E(x_j^2) - E(x_j)^2} = \frac{Cov[x_j, y]}{var[x_j]}$$

$$w_0 = \frac{S(y)\|x_j\|^2 - S(x_j)(x_j^T y)}{n\|x_j\|^2 - S(x_j)^2} = \frac{n^2(E(y)E(x_j^2) - E(x_j)E(x_j y))}{n^2 \, var(x_j)}$$

$$= \frac{E(y)(\underbrace{E(x_j^2) - E(x_j)^2}_{var(x_j)}) + E(y)E(x_j)^2 - E(x_j)E(x_j y)}{var(x_j)} = E(y) + E(x_j)\frac{E(y)E(x_j) - E(x_j y)}{var(x_j)} =$$

$$E(y) + E(x_j)\frac{-cov(x_j, y)}{var(x_j)} = E[y] - w_j E[x_j]$$