Boosting Algo's

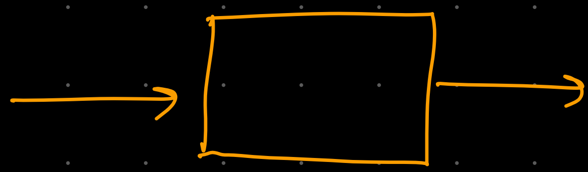
1. AdaBoost
2. Gradient Boost
3. XG Boost

↳ read.csv

EDA

1. SVM/SVR
2. Unsupervised

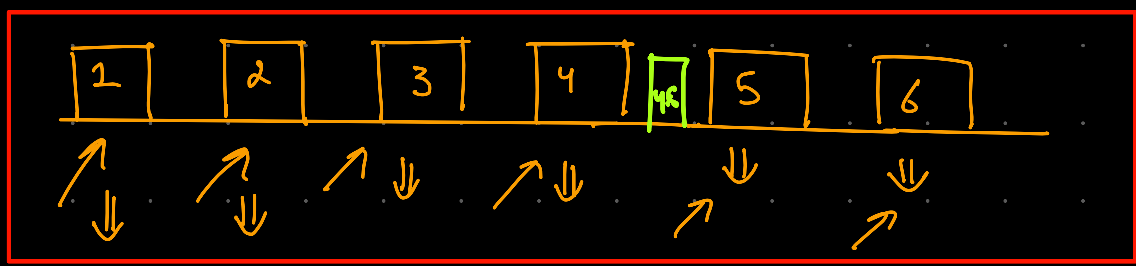
Tomorrow no class

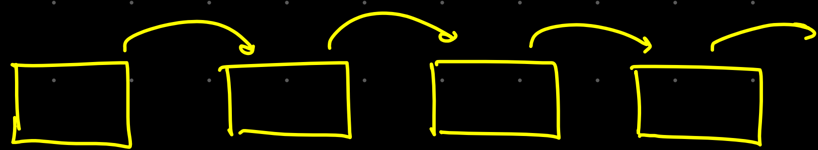


Adaboost

Boosting Algo

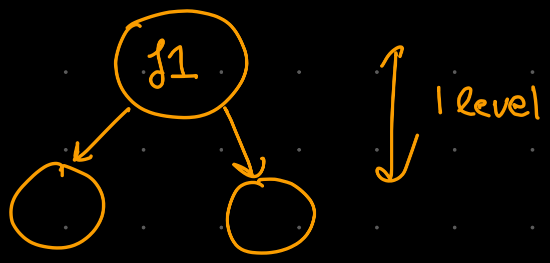
maths.





f1	f2	f3	O/P	wt		New wt
–	–	–	Y	1/7	↓↓	0.058
–	–	–	N	1/7	↓↓	0.058
–	–	–	Y	1/7	← wrong	0.058
–	–	–	N	1/7		$\boxed{0.349}$
–	–	–	N	1/7		0.058
–	–	–	Y	1/7		0.058
–	–	–	Y	1/7		0.058
				1	→	0.697

train a DT with a single depth/level



Stump

1
2
3 → wrong
4
5
6
7

out of 7, 1 I am predicting wrong with my stump after training it.

Total error = $\frac{1}{7}$

Performance of stump = $\frac{1}{2} \ln \left(\frac{1 - TE}{TE} \right)$

$$= \frac{1}{2} \log_e \left(\frac{1 - 1/7}{1/7} \right)$$

$$= \frac{1}{2} \ln \left(\frac{6}{1} \right)$$

$$P_s = 0.895 \quad < 1$$

So now we need to update the weight.

$\Rightarrow$ Correct prediction $\downarrow\downarrow\downarrow$

wrong prediction $\uparrow\uparrow\uparrow$

This will ensure that we send more wrong values to next DT.

New sample weight (correct) $= W_{old} * e^{-P_s}$

$$= \frac{1}{7} * e^{-0.895}$$

$$= 0.058$$

New sample weight for incorrect $= Wt_{old} * e^{P_s}$

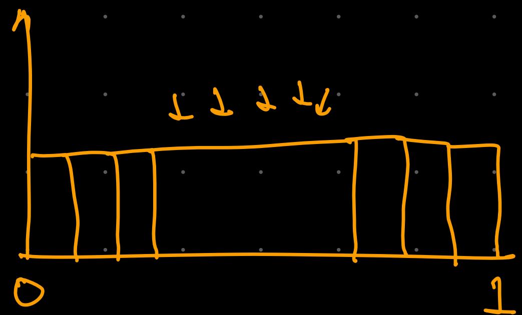
$$= \frac{1}{7} e^{0.895}$$

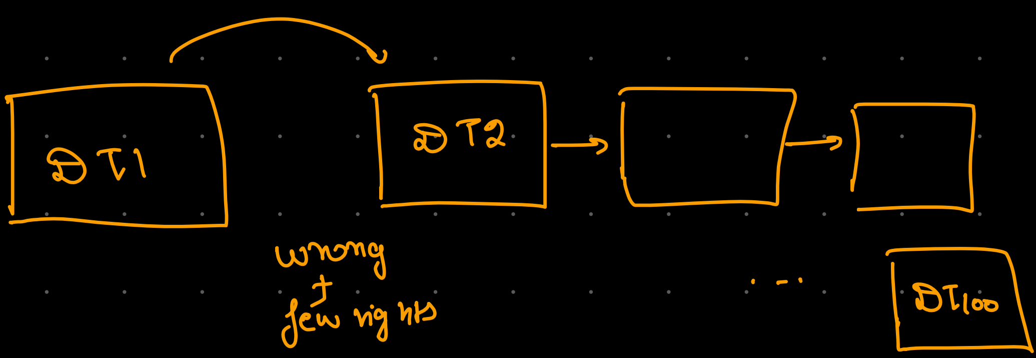
$$= 0.349$$

New wt		Normalized wt	Buckets
0.058	/ 0.697 →	0.083	0 – 0.083
0.058		0.083	0.083 → 0.166
0.058		0.083	0.166 – 0.249
$\boxed{0.349}$		0.500	0.249 – 0.749
0.058		0.083	0.749 –
0.058		0.083	
0.058		0.083	
0.697			

normalized → new wt

1.	0.8742
2.	0.1238
3.	0.5127
4.	0.9873
5.	0.3981
6.	0.7456
7.	0.2319
8.	0.6724
9.	0.0195
10.	0.8902





majority voting = classification

mean = regression.

Q: what is Data Leakage

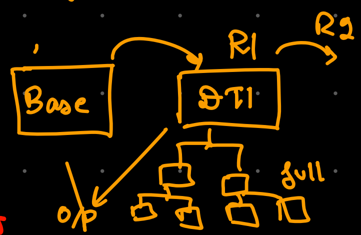
Exam Syllabus $\longleftrightarrow$ Train

$\boxed{\text{Exam Question}}$ $\longleftrightarrow$ $\boxed{\text{Test}}$

Paper leak $\longleftrightarrow$ data leak

Gradient Boosting

Gradient Boosting - a boosting technique



Exp (f_1)	Degree (f_2)	Salary (O/P)	Worst model o/p	error (R1)	(R2)
2	BE	50k	75k	-25k	-23
3	master	70k	75k	-5k	-3
5		80k	75k	+5k	+3
6	PhD	100k	75k	+25k	+23
		300k			

- 2k

3.5 BTech $75 + (-(3)) \approx \boxed{72k}$

① Initialize with a constant value

Base model → O/P which is constant

↳ avg value model which gives out avg value.

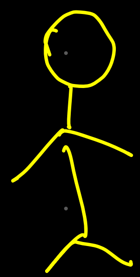
(2) from 1-M calculate pseudo residuals
⇒ error

Now in the next step, we are going to create a DT as exp & education/degree as input features but R1 & not salary as my O/P feature

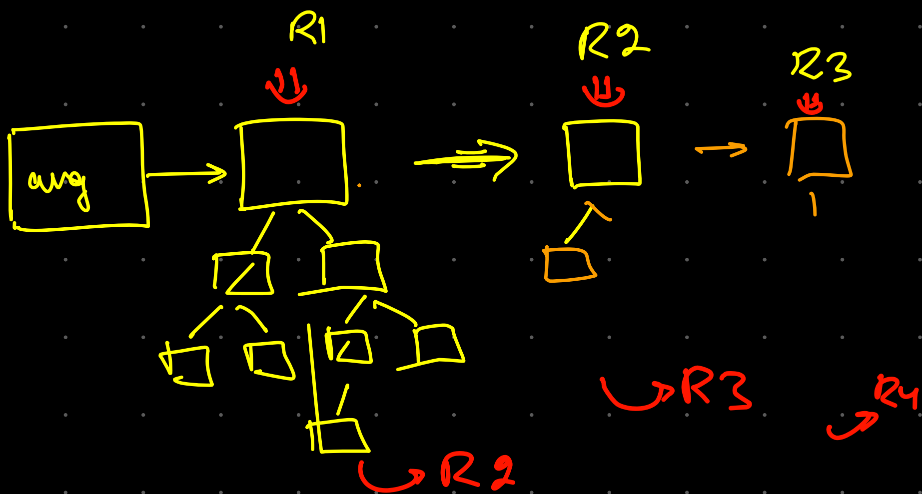
We take R1 as we have to calculate salary & reduce residuals.

Now say O/P of DT1 is R2.

$$70 \; +5 \; -2 \; +10 + 10 - 5$$

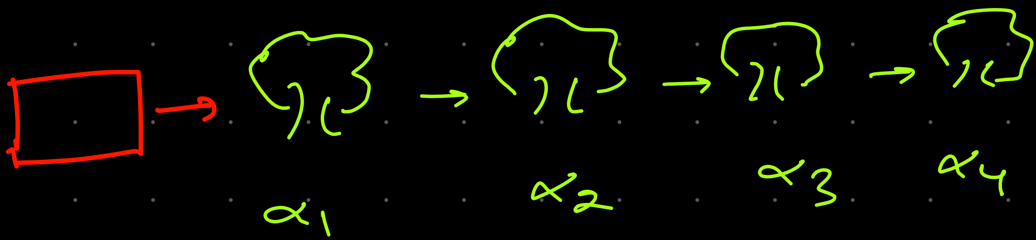


avg.

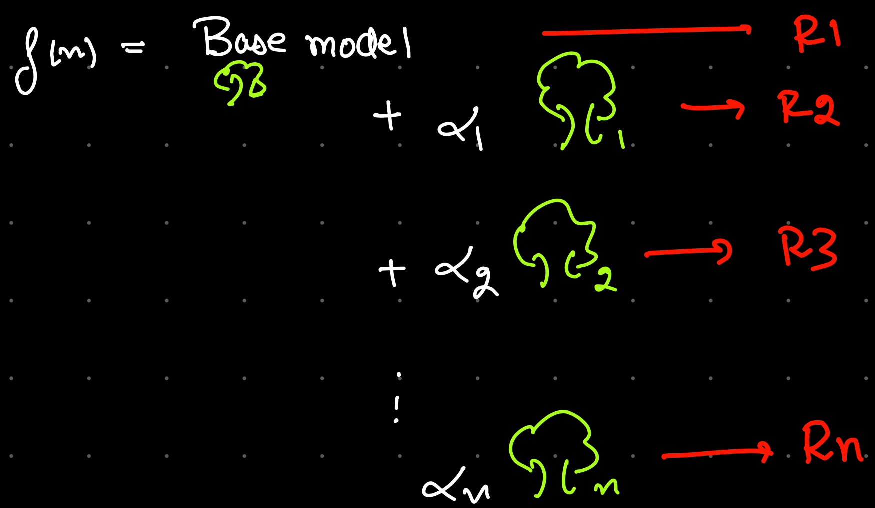


$\alpha_1 \Rightarrow$ learning rate.

$$\Rightarrow \; 75 + \alpha_1(-3)$$

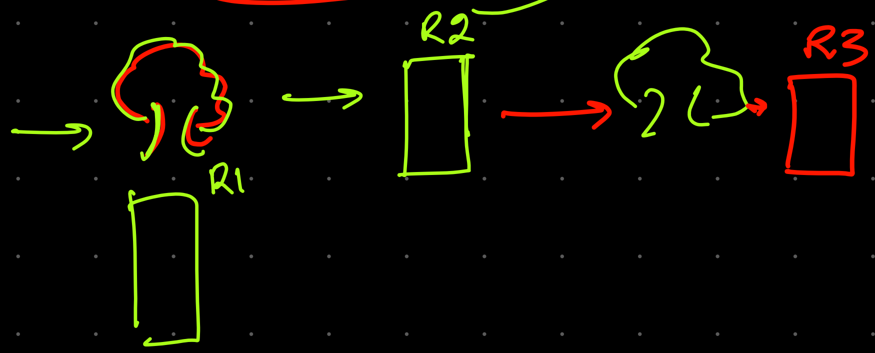


$$f(x) = h_0(x) + \alpha_1 h_1(x_1) + \alpha_2(h_2(x))$$
$$+ \ldots \alpha_n h(x_n)$$



α to reduce overfitting

exp	Degree	R1
3	B.E	-10
4	M.E	5k
5	PhD	3k
6	Bcom	-8k



XG Boost

Salary	Credit	Approval	(0.5) (R1)
<=50	Bad	0	-0.5
<=50	Good	1	0.5
<=50k	Good	1	0.5
>50k	Bad	0	-0.5
>50k	Good	1	0.5
>50k	Normal	1	0.5
<=50k	Normal	0	-0.5

① Construct a tree with root (Base model)

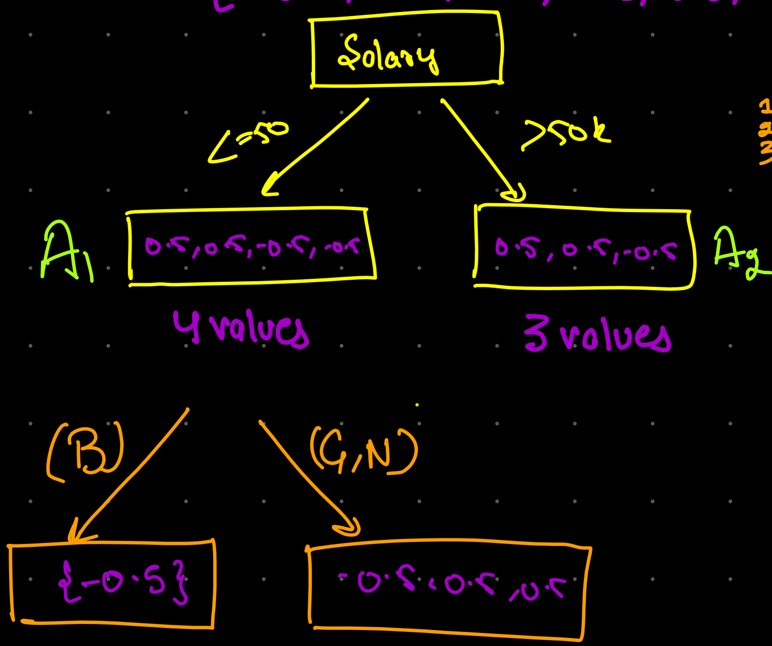
O/P = 0.5

$\boxed{\text{Base}}$

② Construct the next DT.

<=50	Bad	0	-0.5
		1	0.5

$\{-0.5, 0.5, 0.5, 0.5, -0.5, 0.5, -0.5\}$



	Salary	Credit	Approval	(R1)
1	<=50	Bad	0	-0.5
2	<=50	Good	1	0.5
3	<=50k	Good	1	0.5
	>50k	Bad	0	-0.5
	>50k	Good	1	0.5
	>50k	Normal	1	0.5
4	<=50k	Normal	0	-0.5

residuals are the O/P

After making our DT we find similarity wrt.

$$\frac{(\sum \text{Residual})^2}{\sum P_r(1-P_r) + \lambda} \rightarrow \text{hyperparameter}$$

P_r ↙ Probability of base model

$$S.\,Score_{A_V} = \frac{(0)^2}{0.5(1-0.5) + (0.5)(0.5) + (0.5)(0.5) + (0.5)(0.5) + \lambda}$$

$\lambda \rightarrow 0$

$= 0$

$$S.\ Score_{Ag} = \frac{(0.5)^2}{3 * (0.5)(0.5)} + \cancel{x} \rightarrow 0$$

$$= \frac{1}{3}$$

$$S.\ Score_{root} = \frac{(0.5)^2}{(0.5)(0.5) * 7}$$

$$= 0.142$$

$$Gain = \sum Child - root$$

$$= 0 + 0.33 - 0.142$$

$$= 0.188$$

$$\boxed{Gain_{credit} = 0.280}$$

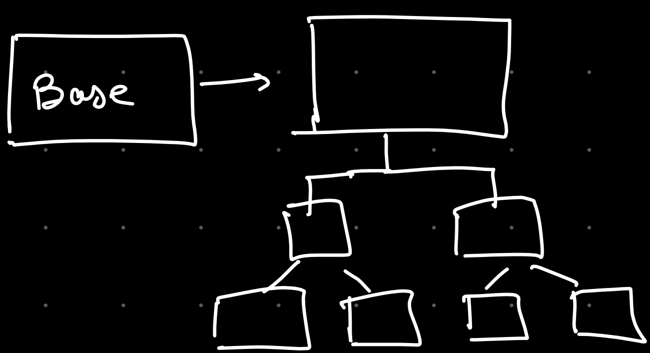
$$S.S_B = \frac{(0.5)^2}{(0.5)^2 + \underbrace{\cancel{\lambda}}_{0}} = 1 \quad \text{(leaf node)}$$

$$S.S_{G,N} = \frac{(0.5)^2}{3 * (0.5)^2} = 0.33$$

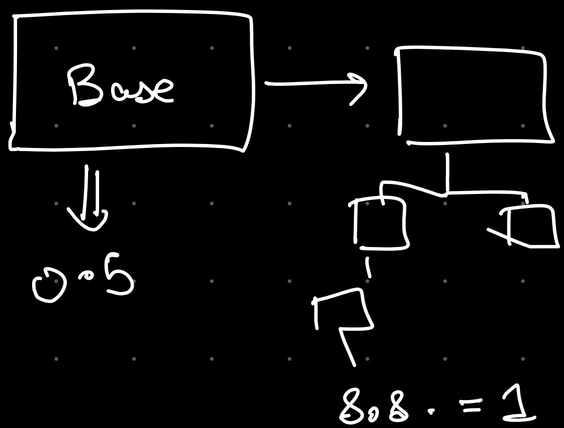
$$S.S._{root} = 0$$

$$\text{Gain} = 1 + 0.3 - 0$$

$$= 1.33$$



<=50 Bad 0 −0.5



$\boxed{\text{Gradient Boosting} — \text{Base model} + \sum \alpha_i \, DT_i}$

In XG. Boost, we apply log loss

$$\log\left(\frac{p}{1-p}\right)$$

for 1^{st} $\Rightarrow \log\left(\frac{0.5}{1-0.5}\right) = \log 1$

$= 0$

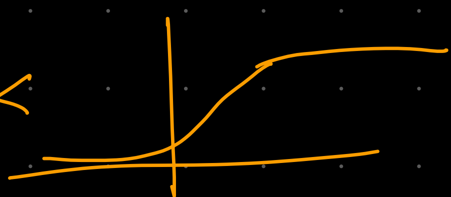
For full D.T., we do below.

$$X_g\text{boost} = \sigma\Big(\underbrace{0}_{\text{log loss of Base model}} + \alpha_1 \big(\underbrace{1}_{\text{S.S / S.W for DT1}}\big)\Big)$$

$$= \sigma\,(0 + 0.5\,(1))$$

$$y = \frac{1}{1+e^{-x}}$$

$\sigma(0.5) \Rightarrow$ classification problem

$0 =<$  < 1

$\sigma(0.5)$

Log loss

$\Uparrow$

$$\text{XgBoost} = \text{Base Model} + \alpha\,(\alpha_1 DT_1 + \alpha_2 DT_2 + \alpha_3 DT_3 \;\vdots\; \alpha_n DT_n)$$

sigmoid

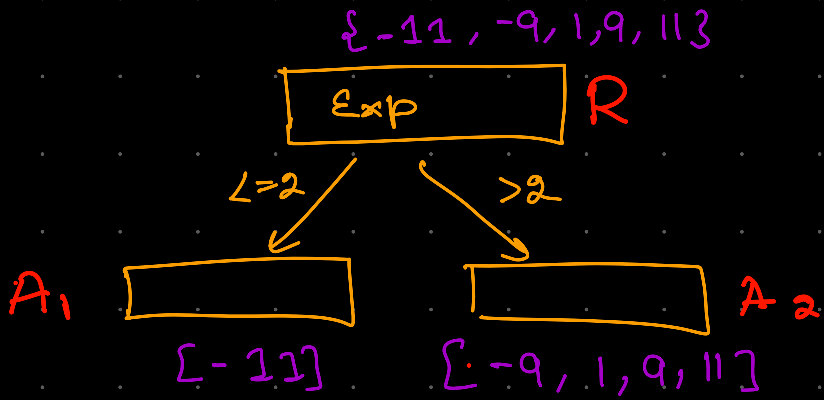
XG Boost Regressor

Exp	IIT	Salary	R1
2	Y	40k	-11
2.5	Y	42k	-9
3	N	52k	1
4	N	60k	9
4.5	Y	62k	11

1. Base model

avg of all salary

② Mak DT with Exp, Gap and output as R1



$$S.W.\ (\text{regression}) = \frac{(\Sigma(\text{residual}))^2}{\text{No. of residuals} + \lambda}$$

$$S.W_{root} = \frac{1^2}{5+1} = \frac{1}{6} = 0.166$$

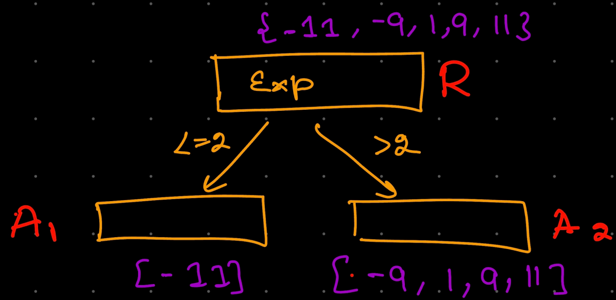
$$S.W_{A_1} = \frac{(-11)^2}{1+1} = \frac{121}{2} = 605$$

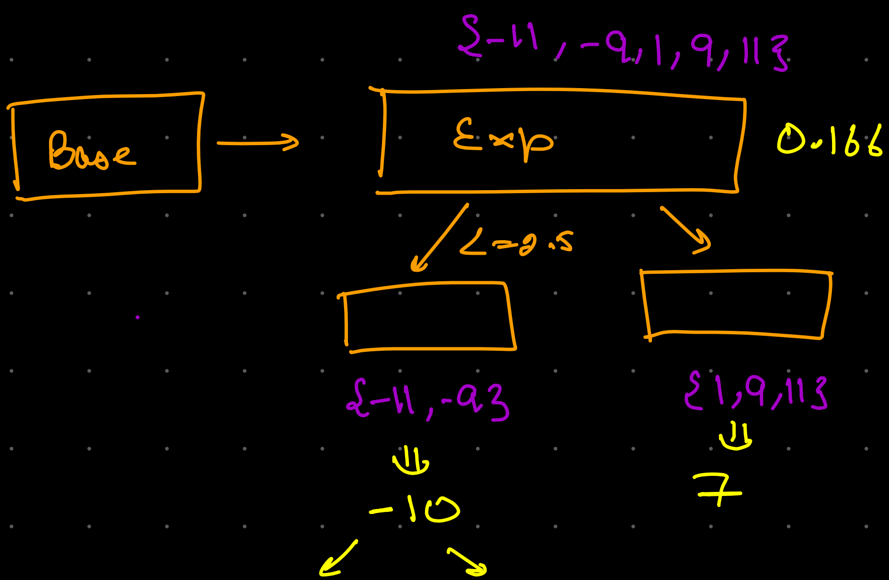
$$S.W._{A_2} = \frac{(12)^2}{4+1} = \frac{144}{5} = 28.8$$

$$\text{Gain} = \Sigma\,\text{child} - \Sigma\,\text{root}$$

$$= 60.5 + 28.8 - 0.166$$

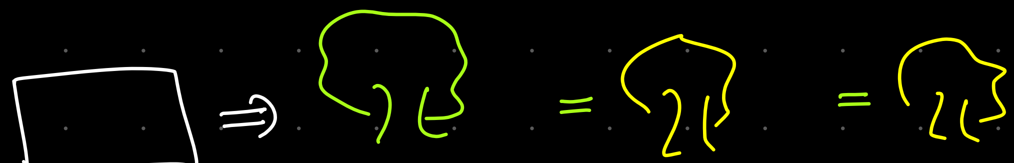
$$= 89.13$$





2Y of exp.

$$= 51 + \alpha_1(-10)$$



$$\Downarrow \qquad\qquad \Downarrow$$

$$51 \qquad\qquad -10$$

$$51 + \alpha_1(-10)$$

Base model $+ \alpha_1 DT_1$
$+ \alpha_2 DT_2 \; \ldots$
$\alpha_n DT_n$