Underfitting
&
Overfitting

ex: Sample size = 200

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
(160) (40) (160) (40)

model.fit(x_train, y_train)

y_pred = model.predict(x_test) # 40

Test data = r2_score(y_test, y_pred)
= 0.75

Train data

y_train_pred = model.predict(x_train) = 160

Train data A = r^2(sc (y train-pred, y train))

= 0.88

1. Traind A = 0.88 → 0.13
2. Test-d A = 0.75

Training data A is high (0.58 %.)

Test data A is low (0.75 %)

∴ Overfitting model

Train data A is low 0.75 %

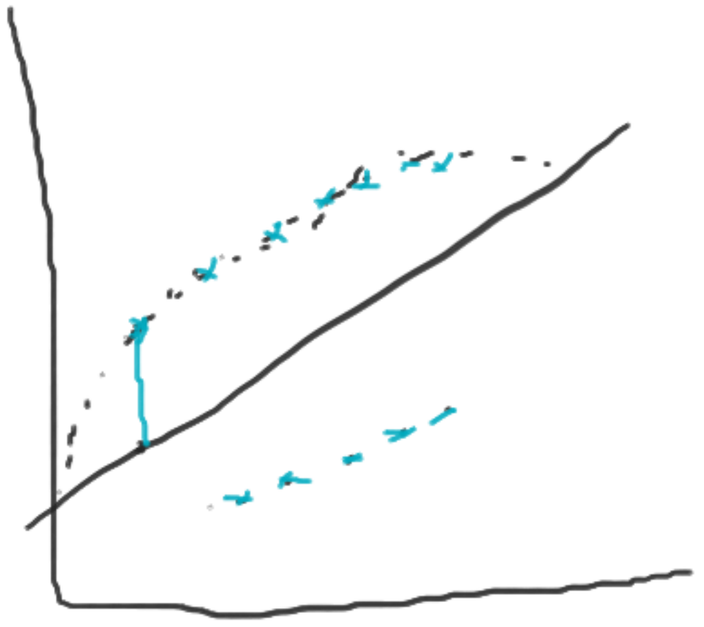
Test data A is low 0.75 %

∴ Underfitting model

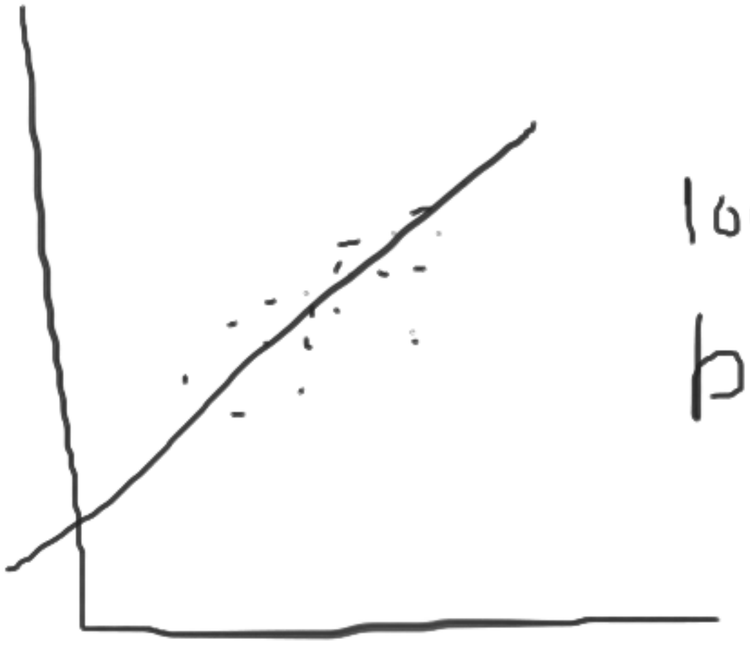
Bias = diff betw avg of pred

Variance = How much scattered are the predicted from their
actual value.

Bias

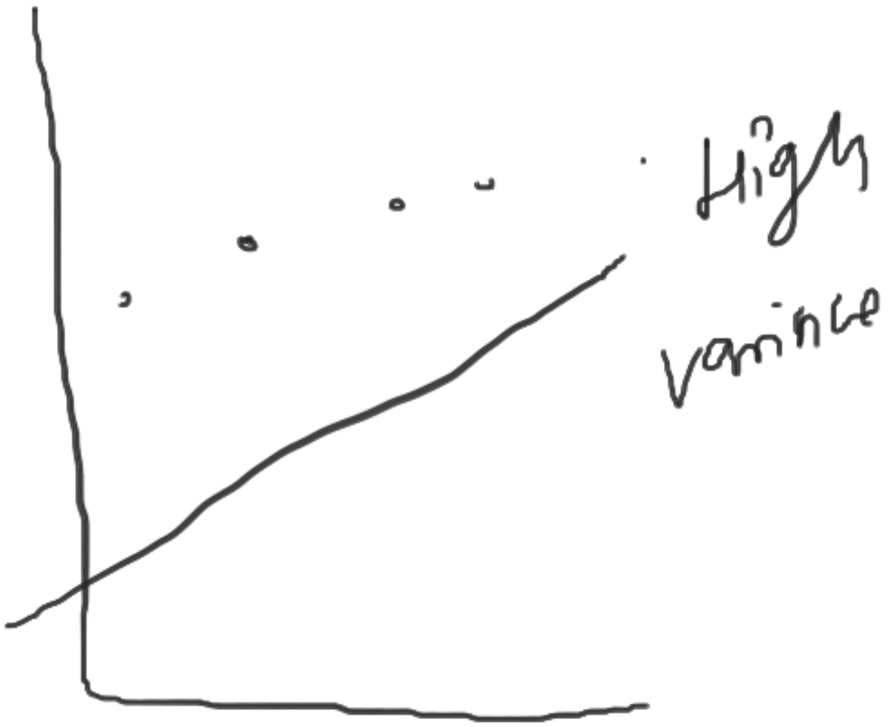


High bias



low bias

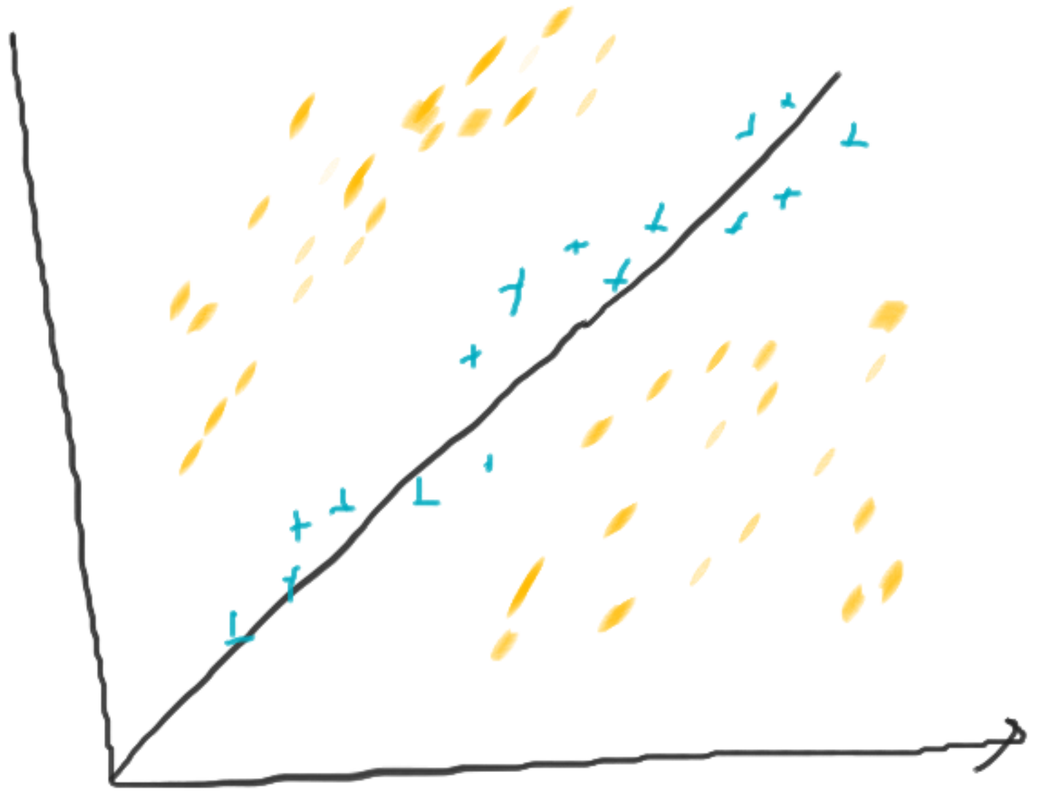
Variance.



High variance

low variance

Overfitting

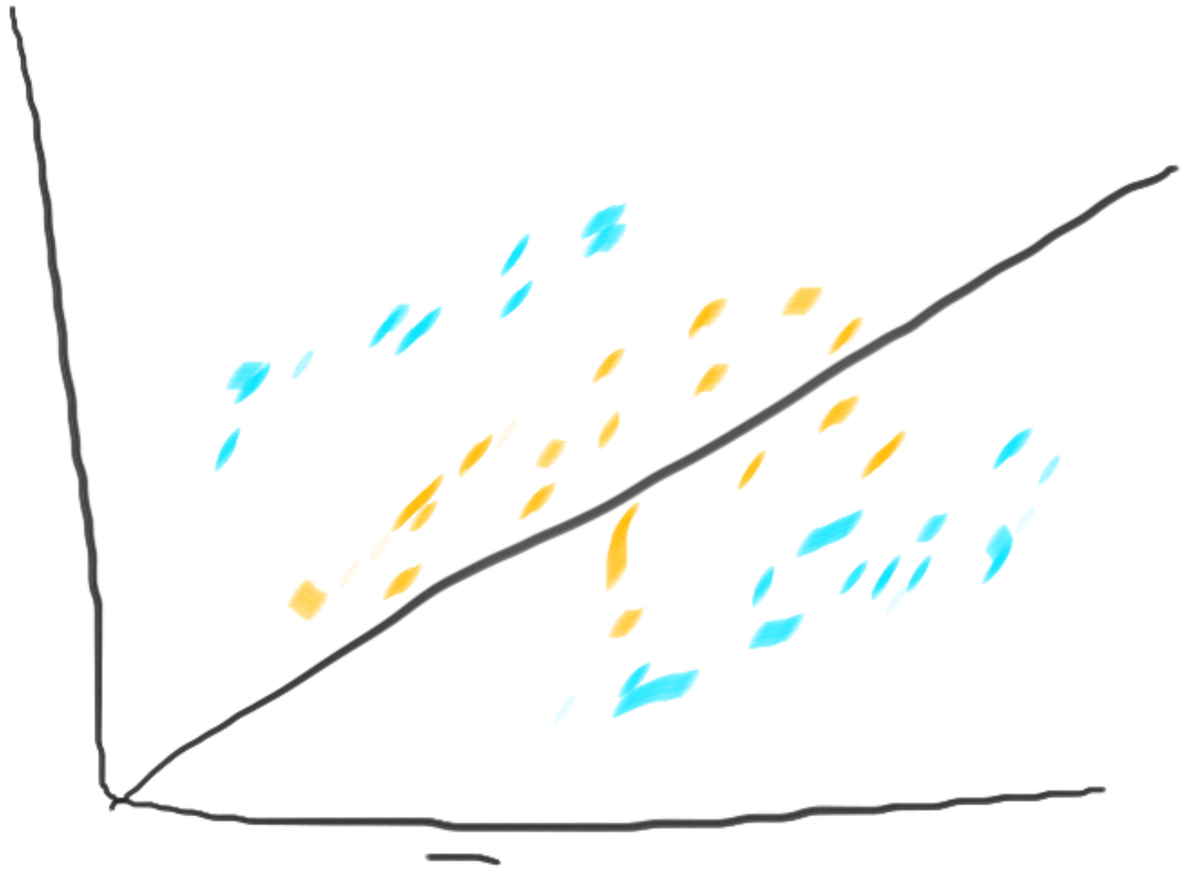


test =

train =

low bias
high variance

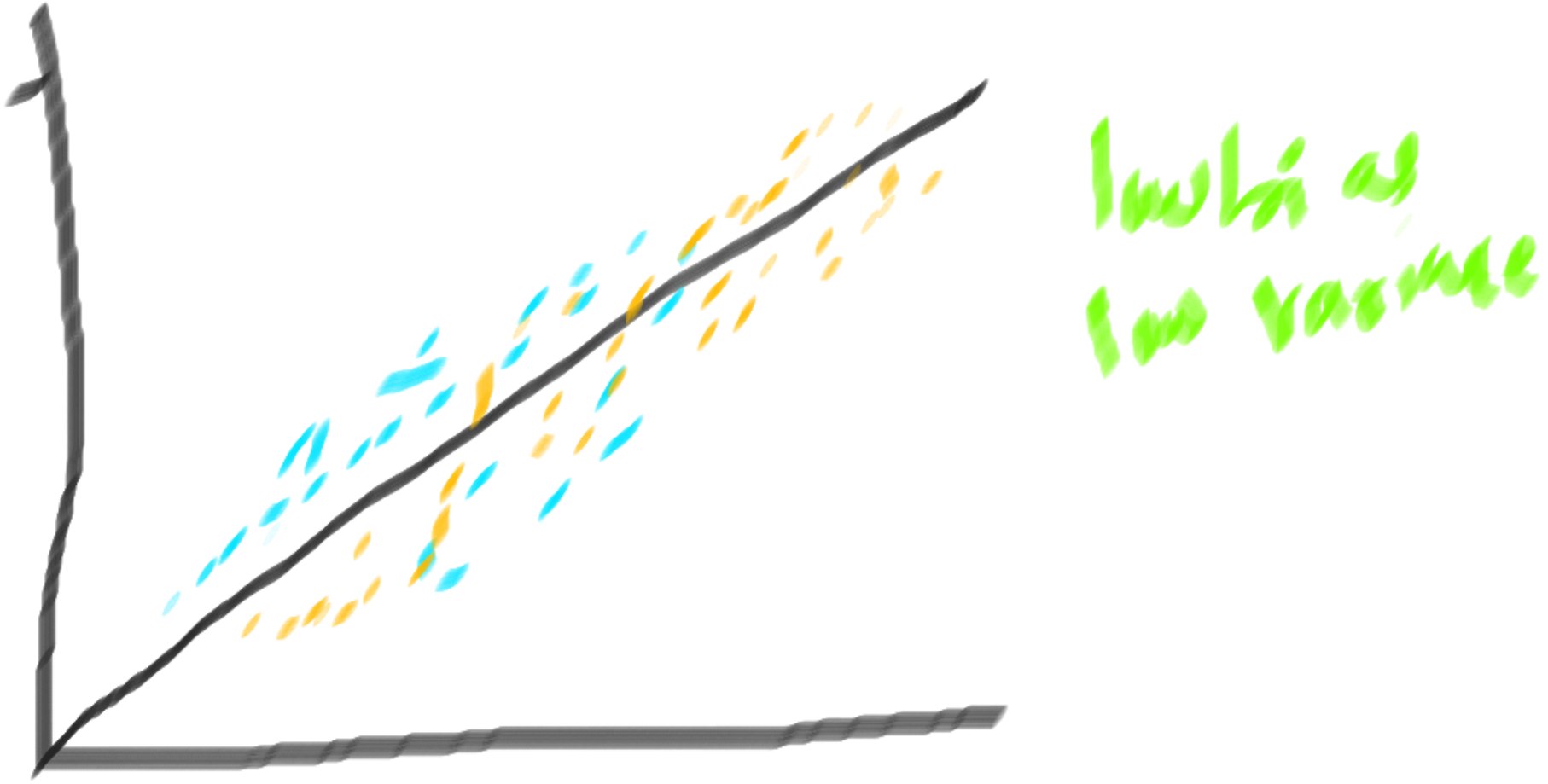
Underfitting



Test : ... → Var
Train : ... → Bias

Low variance
High bias

Log model



High bias ⇒ Less accuracy on train

Low bias ⇒ Good accuracy

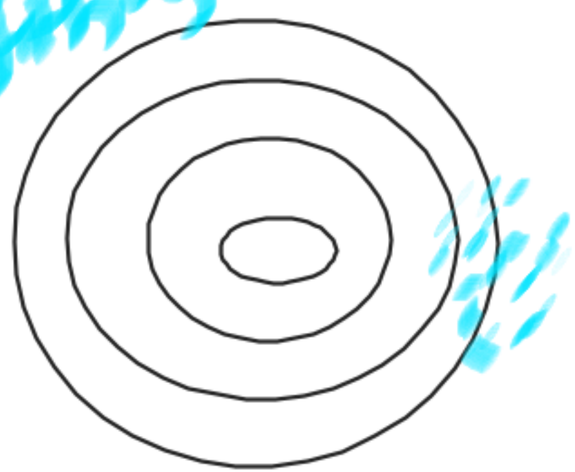
Low variance ⇒ Good accuracy

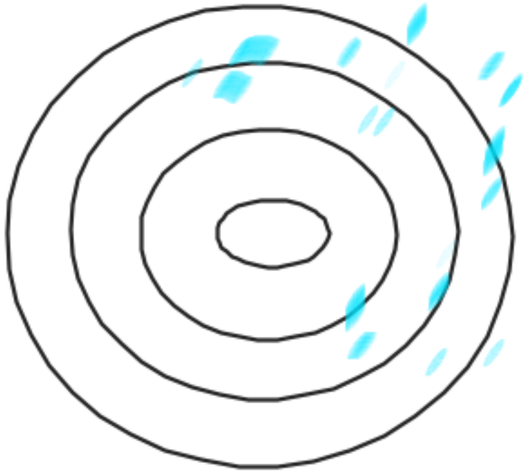
High variance ⇒ Less accuracy

Low variance

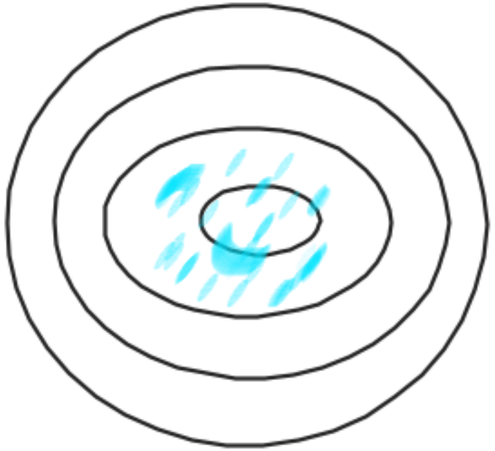
High variance

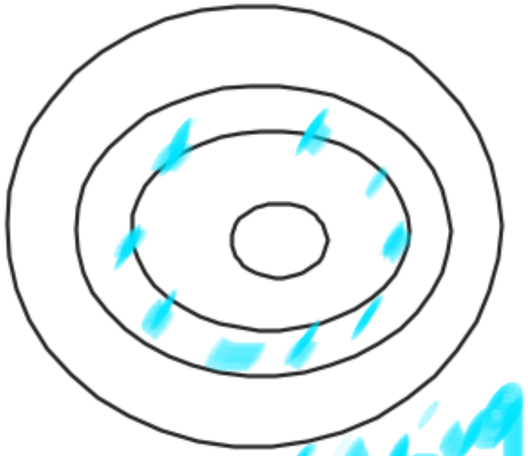
High bias

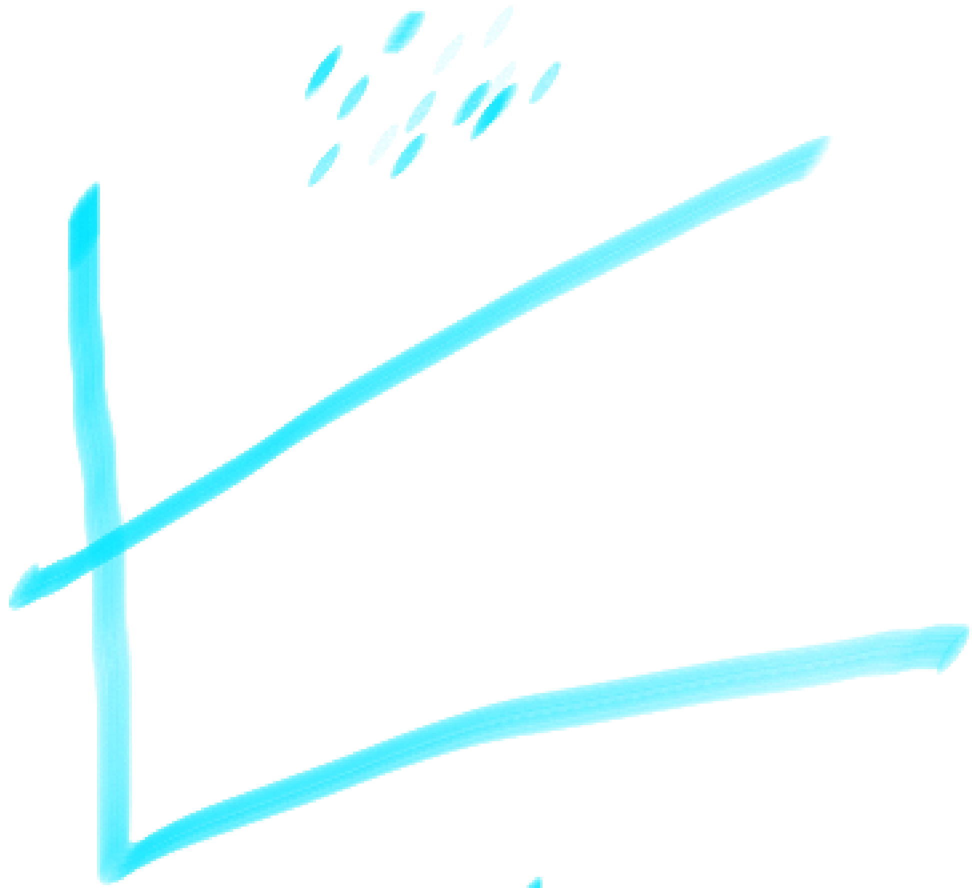




Low bias







Build a model anyways

65 to 70

~ 73-83

93.9

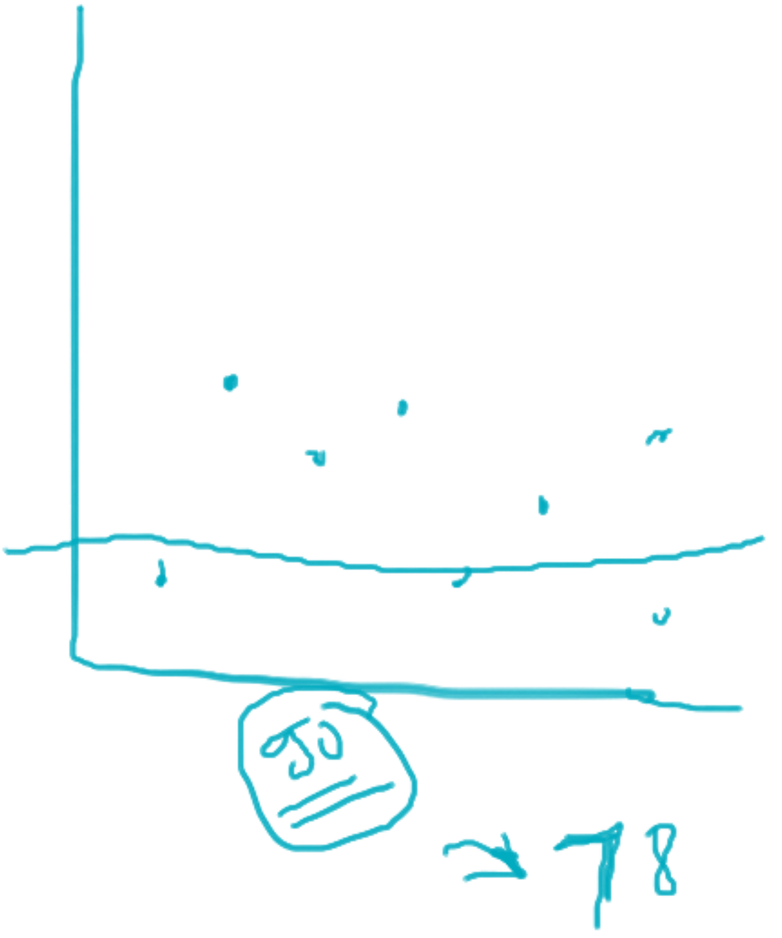

error
test error
training error
model complexity

	Age	Ht	wt	Bmi	
Chetan	40	179	60	24	yes

10

41	179	65	25	yes


overfitting
96
Ben.


50
78