* Agenda and Recap
    * Normality of Residuals
    * Homoscedasticity
    * Auto-collinearity

  3

* Gradient Descent Variants

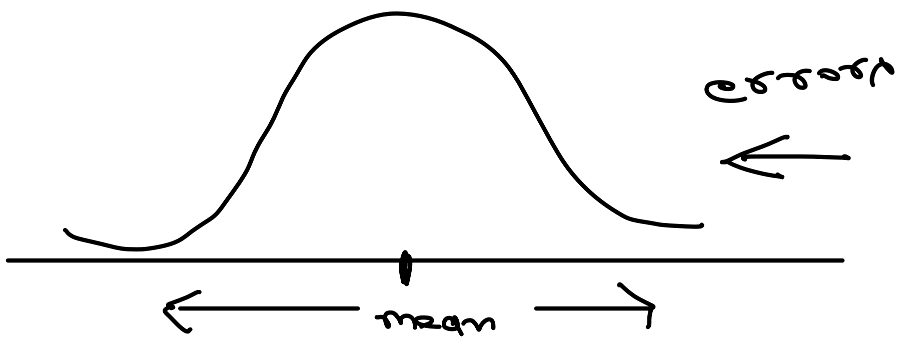
* Polynomial Regression

* Generalization and Occam's razor

* Underfitting and overfitting

* Bias Variance Tradeoff

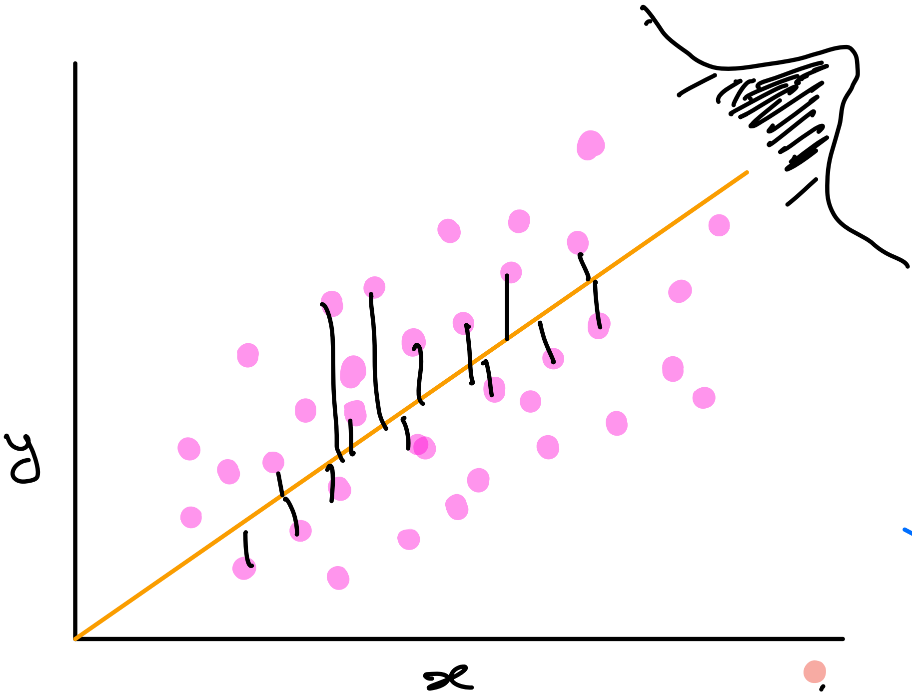
Assumption 3: Errors are Normally distributed

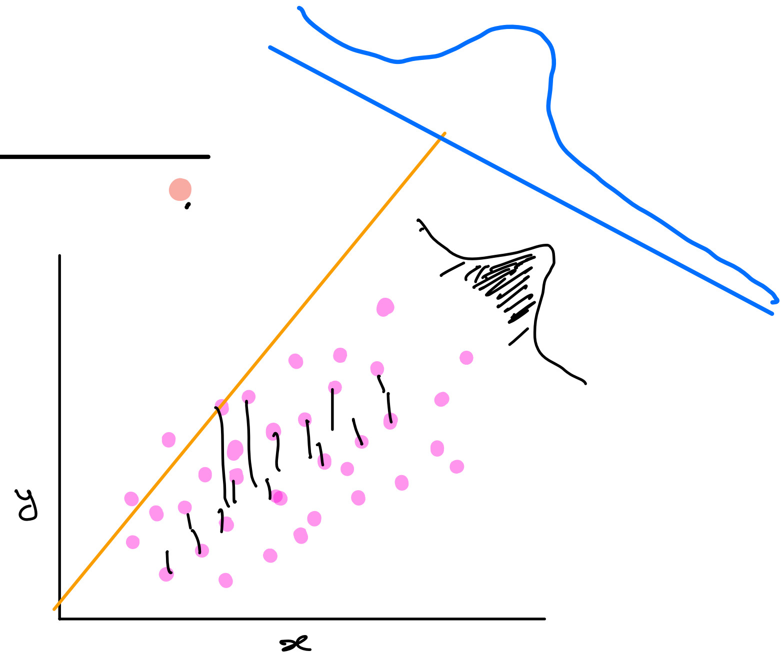


Step 1: Build Model

Step 2: Calculate Error

Step 3: Plot Errors with Histogram



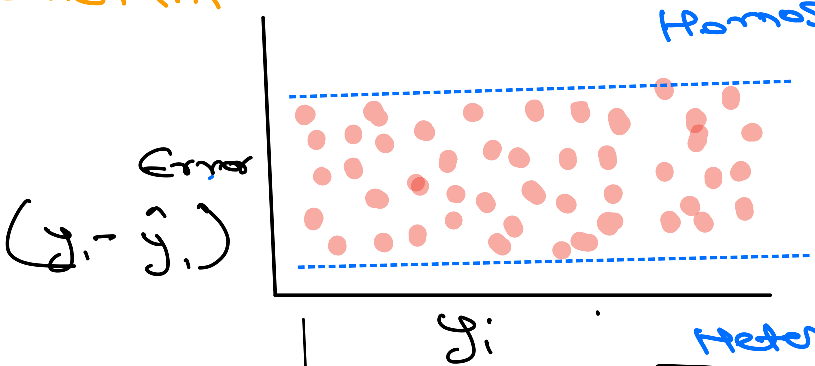


Assumption 4 : Heteroskedasticity should not exist

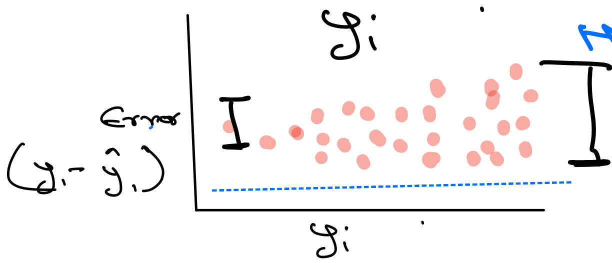
⇒ Variance of residuals (ϵ_i) vs predictions ($\hat{y}_i$) should be constant

$\boxed{y_i \text{ vs } \epsilon_i}$

Homoscedasticity



Heteroscedasticity



* Goldfeld Quandt Test

⇒ Null Hypothesis: Dataset has Homoscedasticity

* If p value $\leq$ significant level Threshold

⇒ reject Null Hypothesis

* To mitigate:

⇒ Remove outlier

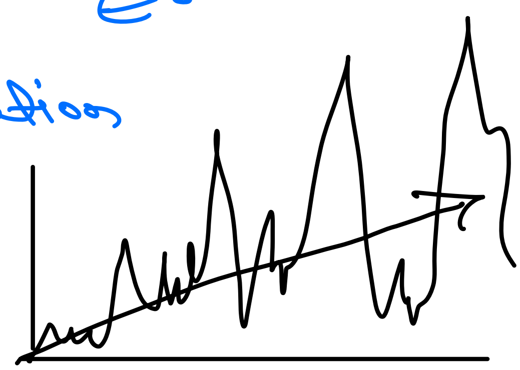
⇒ Perform Non-Linear Transformation such as box-cox

(Correlate between Errors)

Assumption 5: No Autocorrelation

↓

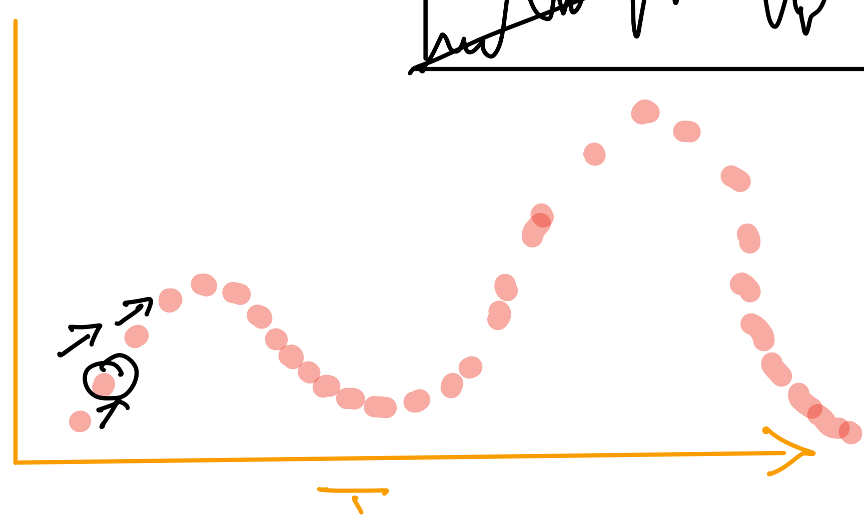
Self



$\boxed{\epsilon_i \text{ vs } T}$

⇒ Time-series Model

ϵ_i



Variants of SGD

(Just 1 update per iteration)

$$w_j^{new} = w_j^{old} - \eta \frac{\partial L}{\partial w_j^{old}}$$

1000 iter ⇒ 1000 updates

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

for One Single iteration/update

$\boxed{\text{1 Million}}$

⇒ 1 million Errors and Gradients

⇒ Update Weights

1 million

K ⇒ between 1 and N

let's K = 256

$$\sum_{i=1}^{\text{Batches}} \frac{2}{K} \sum_{i=1}^{K} (\hat{y}_i - y_i) \times x_{ij}$$

$$w_j^{new} = w_j^{old} - \eta \frac{\partial L}{\partial w_j^{old}}$$

Batches ⇒ $\frac{1 \text{ million}}{256}$ Updates ←

(updates per iteration)

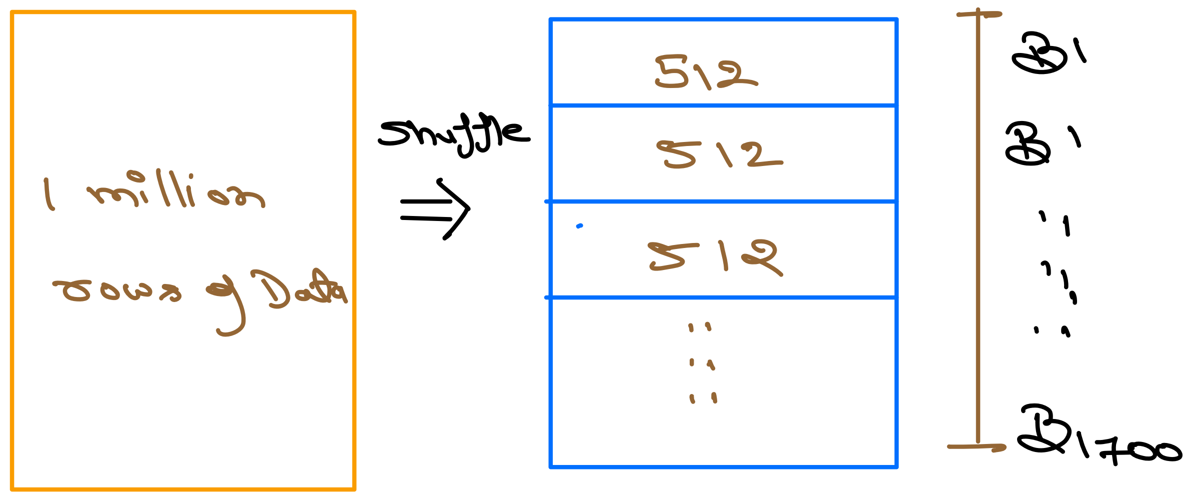
1000 iters → 1000 × Batches

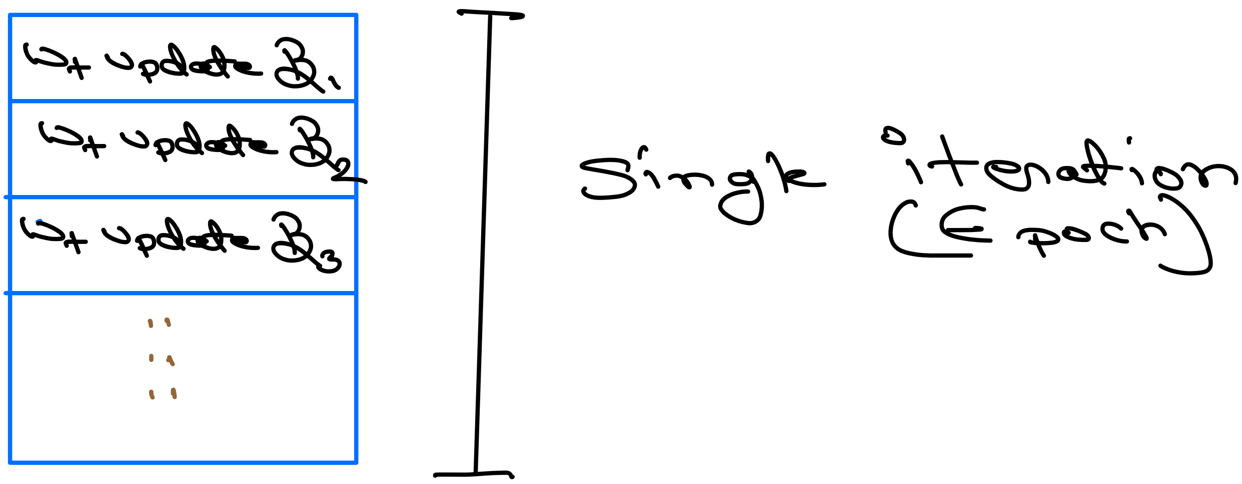
K → 1 (Stochastic Gradient Descent (SGD))

→ N (Batch GD)

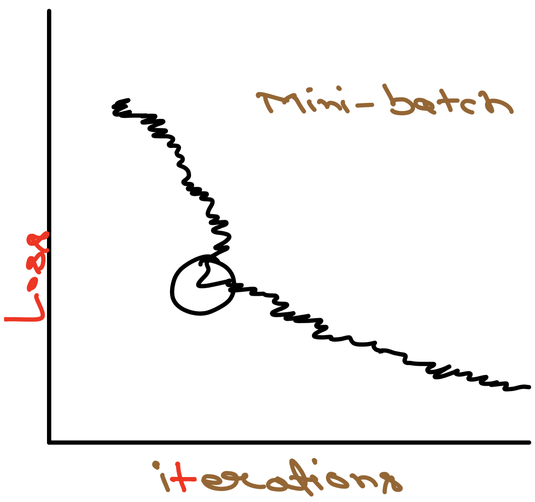
→ (1, N) (Mini-Batch GD) MGD

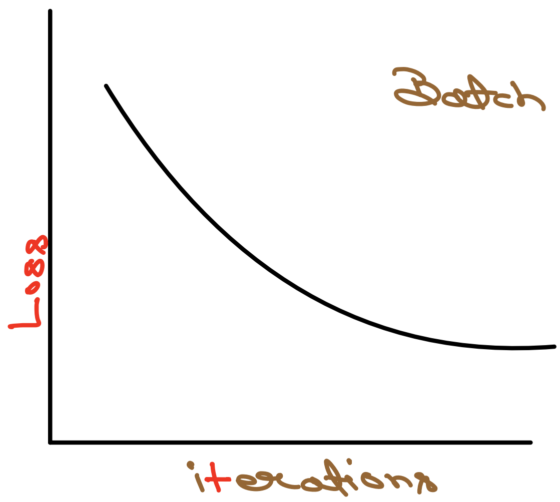
(Weight update in Single Epoch)





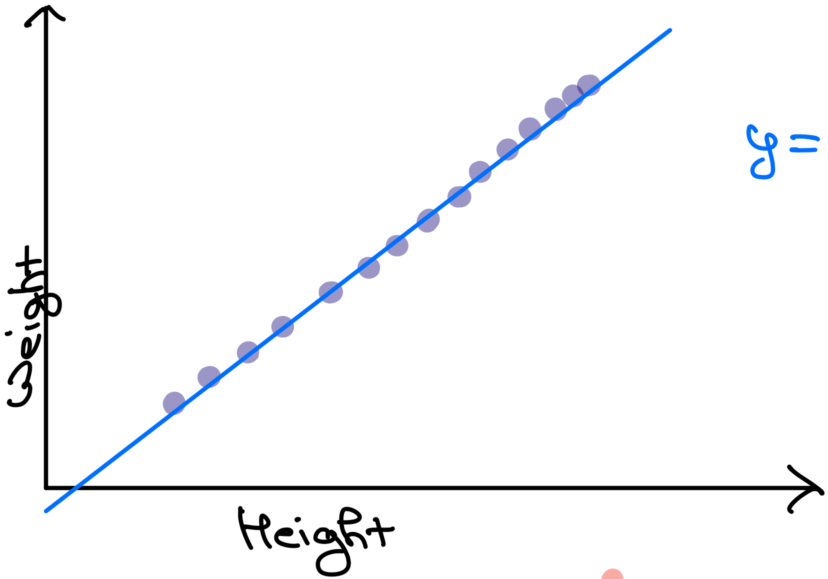
Comparison : Minibatch vs Batch



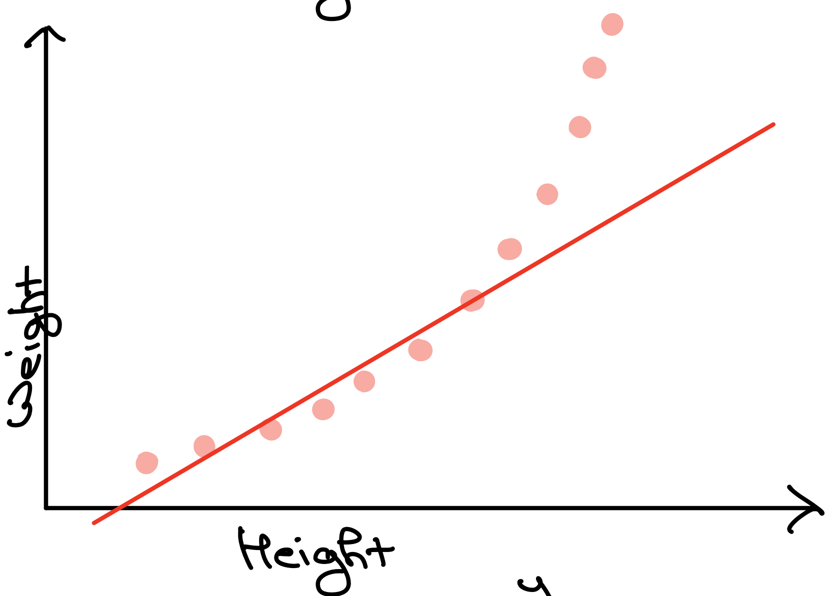


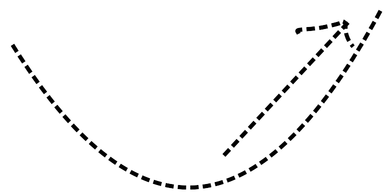
Polynomial Regression

John



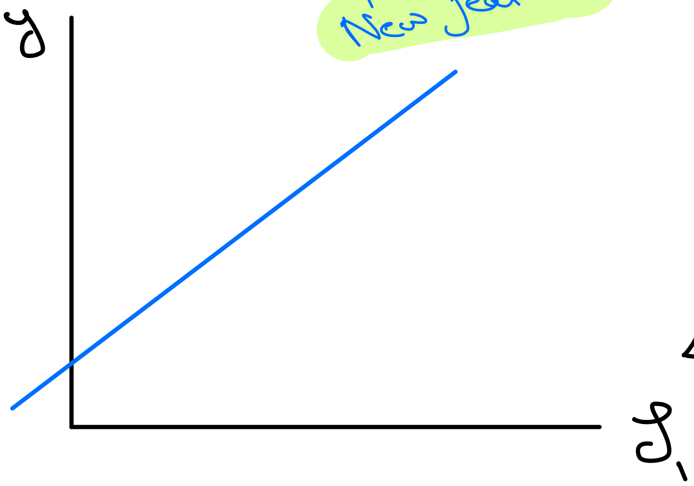
$y = \omega_1 x_1 + \omega_0$

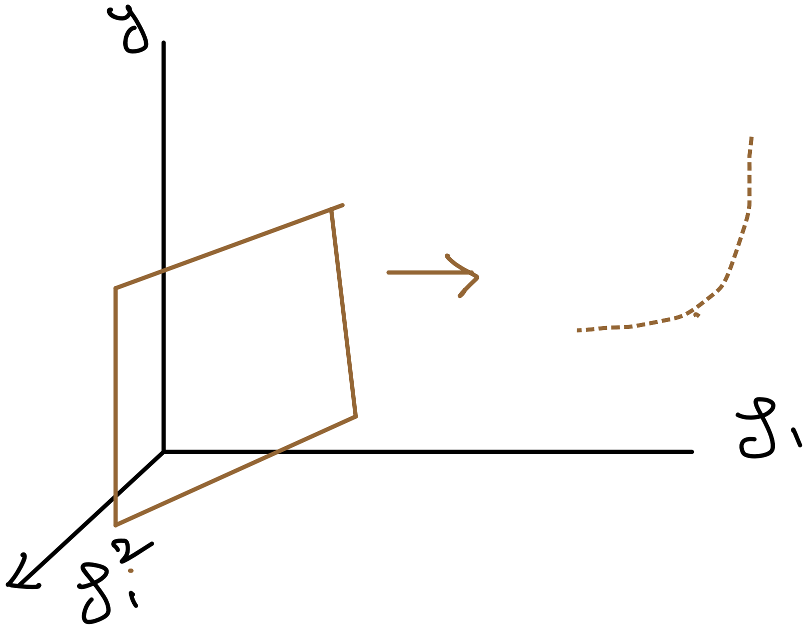




$y = \omega_1 f_1 + \omega_2 f_1^2 + \omega_0$

New feature





* How Does this Look

→ In data

$f_2 \Rightarrow f_2^2$

$f_1 \times f_2$

x	y
x_{11}	y_1
x_{21}	y_2
x_{31}	y_3
x_{41}	y_4
x_{n1}	y_n

$\Rightarrow$

f_1^2	f_1	
x^2	x_1	y
x_{11}^2	x_{11}	y_1
x_{21}^2	x_{21}	y_2
x_{31}^2	x_{31}	y_3
⋮	x_{41}	y_4
x_{n1}^2	x_{n1}	y_n

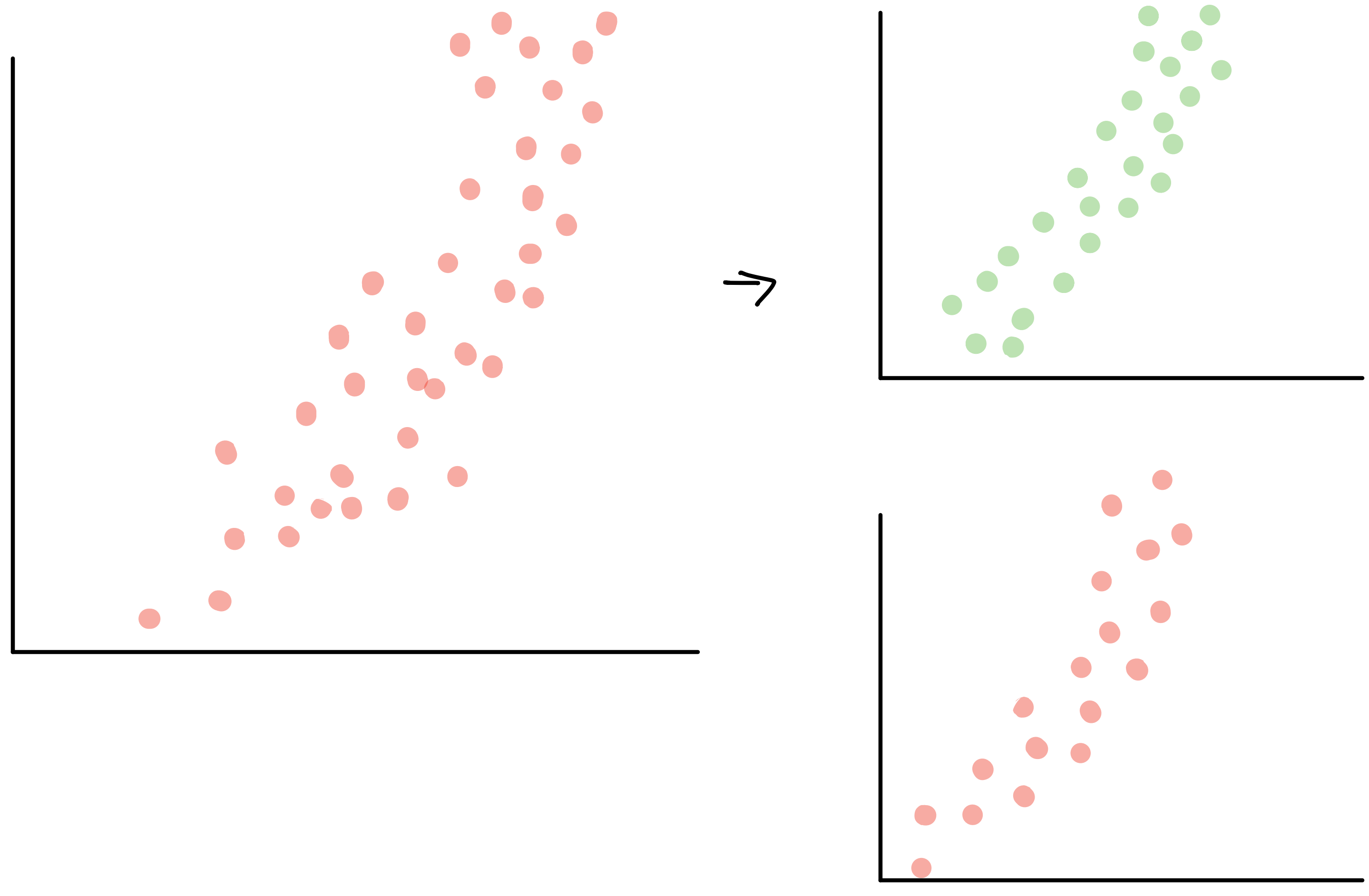
→ In plot

* What about Multi-collinearity?

$f_2 = \alpha f_1 + \beta \leftarrow$ Linear Relationship

$f_2 = f_1 \times f_2 \leftarrow$ Non Linear

Generalization and Occam's Razor

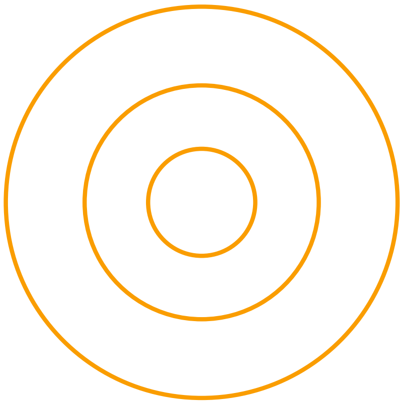


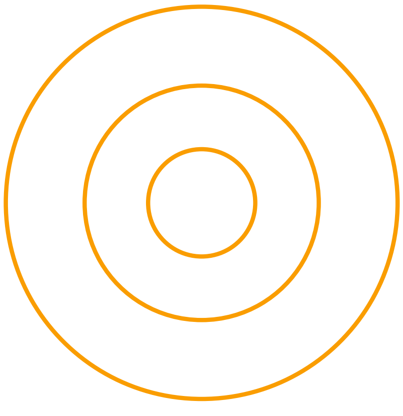
Simplest

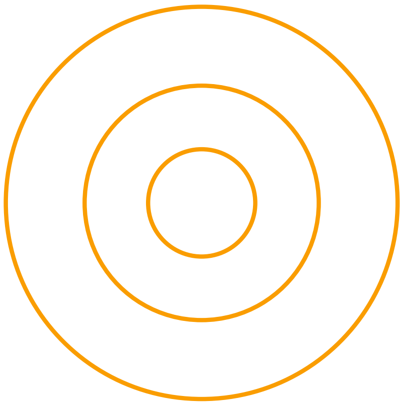
Model	Training	Testing

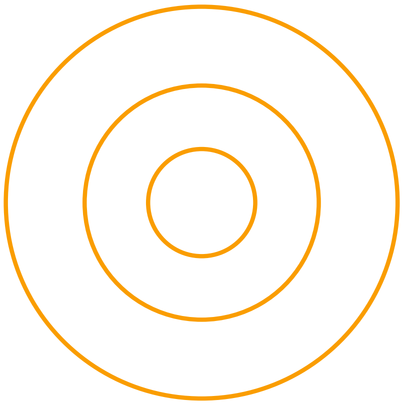
Complex

Bias and Variance Trade-off









Bias vs Variance