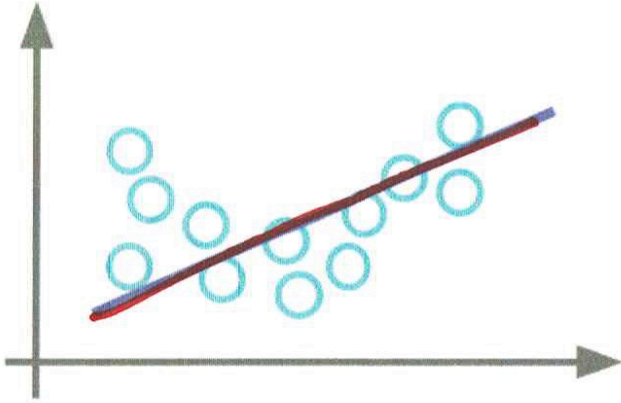
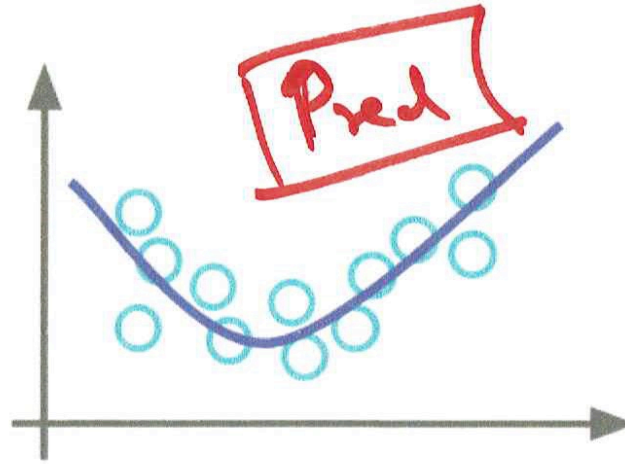


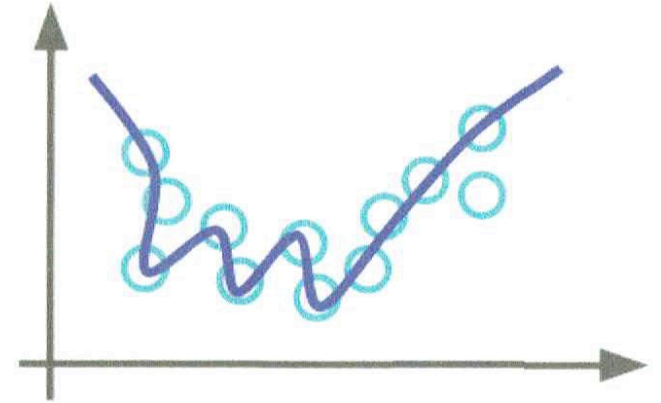
$$\text{Data} = \underbrace{\text{Inf}}_{\text{model}} + \text{Noise}$$



Under



✓



Over

$L_1 \propto L_2$  Reg

$L_1 \rightarrow$  Lasso  $\leftarrow$

$L_2 \rightarrow$  Ridge

Min  $\downarrow$

$\uparrow$

$L(y, \hat{y}) + \lambda(\text{penalty})$

$\rightarrow$   $\sum |w_i|$  Lasso

$\rightarrow$   $\sum (w_i)^2$  Ridge

$$\omega^j = \omega^{j-1} - \eta \underline{\nabla_{\omega} l(\omega)}$$

SGD  
with momentum

$$\underline{\omega^j = \omega^{j-1} - \eta g^j}$$

$$g^j = \nabla_{\omega} l(\omega)$$

add  
momentum

$$\omega^j = \omega^{j-1} - \eta \underline{(\alpha g^j + (1-\alpha) g^{j-1})}$$