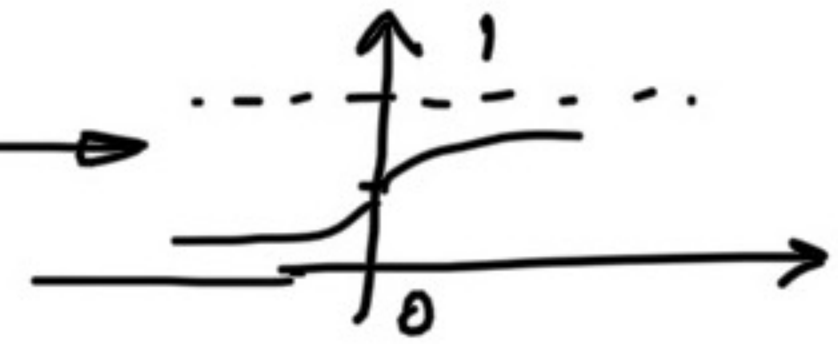
Regularization

example : $\hat{y} = \sigma(w x_1 + w x_2 + b)$ prediction

if $w = 1$ $\longrightarrow$ 

if $w = 10$ $\longrightarrow$ 

too certain

large coefficients

solution

In regularization, we want to penalize <u>large</u> weights.

good for feature selection $\leftarrow$ $\text{Error} + \lambda(|w_1| + \cdots + |w_n|)$ l1 - regularization

usually good for traing models $\leftarrow$ $\text{Error} + \lambda(w_1^2 + \cdots + w_n^2)$ l2 - regularization