

$|\beta|_2^2$: 16, Trainable: last layer

