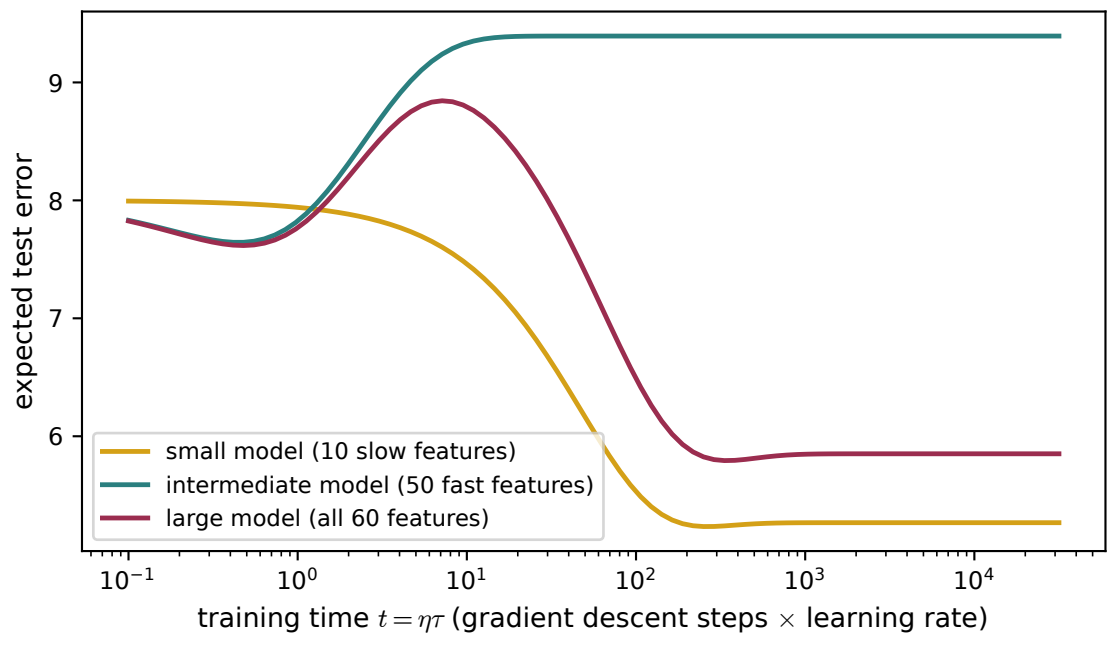

expected test error
9
8
7
6
small model (10 slow features)
intermediate model (50 fast features)
large model (all 60 features)
$10^{-1}$
$10^{0}$
$10^{1}$
$10^{2}$
$10^{3}$
$10^{4}$
training time $t = \eta\tau$ (gradient descent steps $\times$ learning rate)