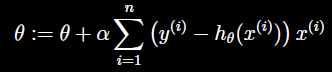
* Gradient descent (batch, stochastic, and mini-batch)
  + CS229 notes (9-13)



* + Often, stochastic gradient descent gets theta close to the minimum much faster than batch gradient descent (Note: however, it may never converge to the minimum, and the parameters theta will keep oscillating around the minimum of J(theta); but in practice most of the values near the minimum will be reasonably good approximations to the true minimum, therefore, when training set is large, stochastic is often preferred over batch)
* Normal equation
  + CS229 notes (13-15)



* + 