Homework Solutions, Section 9.9

1. For the line of best fit,

   (a)
   $$E(m,b) = (0-(m(-1)+b))^2 + (1-(m+b))^2 + (3-(2m+b))^2 + (2-(3m+b))^2$$

   (b) Computing the partial derivatives and simplifying, we get:
   $$\frac{\partial E}{\partial m} = 30m + 10b - 26, \qquad \frac{\partial E}{\partial b} = 10m + 8b - 12$$

   (c) At the point $(0,-1)$, the gradient becomes $\langle -36, -20\rangle$. Therefore, our update gives:
   $$\begin{bmatrix} 0 \\ -1 \end{bmatrix} - \frac{1}{10}\begin{bmatrix} -36 \\ -20 \end{bmatrix} = \begin{bmatrix} 3.6 \\ 1 \end{bmatrix}$$

   (d) For stochastic gradient descent, we are told to use data point 3 for the estimation:
   $$\nabla E = \langle -12 + 8m + 4b, -6 + 4m + 2b\rangle|_{(0,-1)} = \langle -16, -8\rangle$$
   so that the update rule is now:
   $$\begin{bmatrix} 0 \\ -1 \end{bmatrix} - \frac{1}{10}\begin{bmatrix} -16 \\ -8 \end{bmatrix} = \begin{bmatrix} 1.6 \\ -0.2 \end{bmatrix}$$

2. (Video shows the solution on Octave-online)

   In my run on Octave-online, I found that regular gradient descent with scaled data and fixed $\alpha = 0.01$ converged to our solution in 16 steps with an error of 13.31 (units here wouldn't make sense unless we scaled them back out). The predicted price of $2,000$ square feet home was about \$419,500.00. This was high due to a few very expensive homes shifting the mean up.

   Using the adaptive step size, we converged in 13 steps.

   Using stochastic gradient descent, we didn't converge in the usual sense, but after 350 iterations, we came close to the minimum (about 14).

   Side Remark: In the "real world", we would not use stochastic gradient descent for this problem (it should be reserved for problems with a lot of data), but it is useful to see how it performs for "easier" problems.