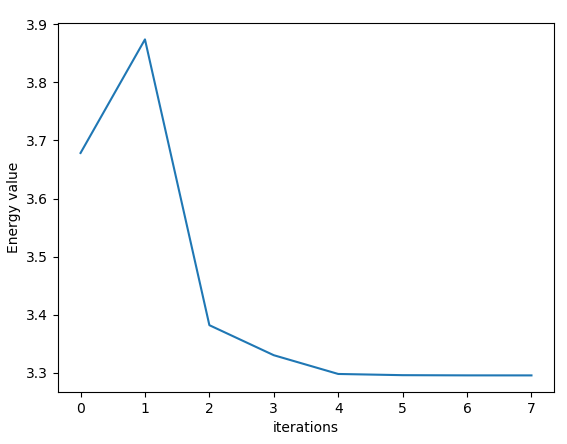
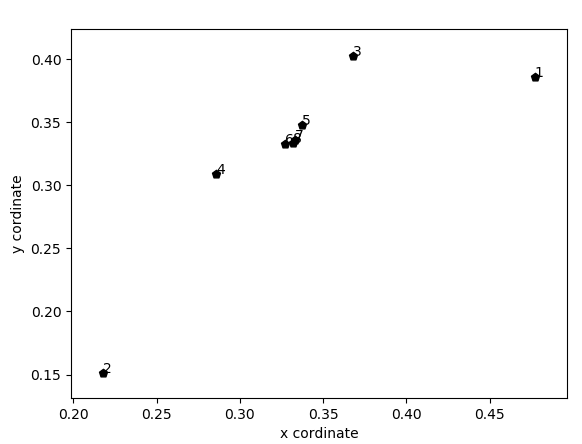
Q2 b)

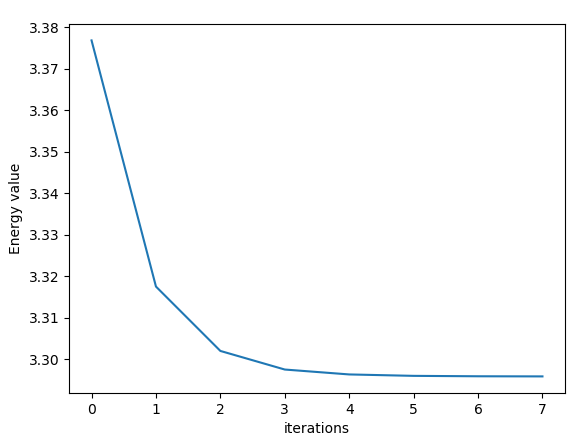
Initial wo = [0.28332678,0.57361592]

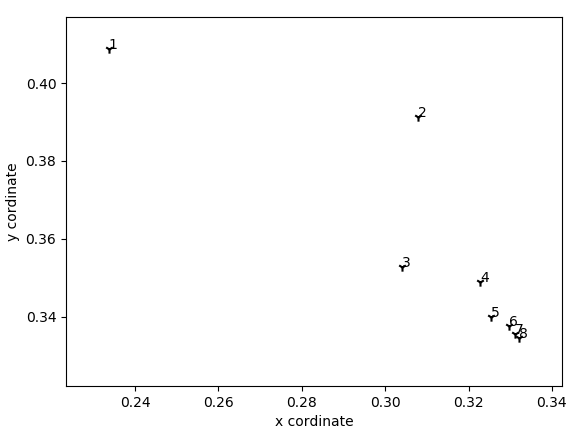
learning rate = 0.05



c) Initial wo = [0.28332678,0.57361592]

learning rate = 0.05





d) Newton method converges slightly better than the gradient descent. As we can see in the above example that Newton method converges within approximately 6 iteration and gradient descent converges 7 iterations. As the function that we were trying to converge had only small domain so both methods perform almost similarly. Newton method converges slightly faster than the gradient descent. For some initial weights (Eg [0.19333942, 0.13318579]) where gradient descent converged, Newton method was not converging at all even with same value of learning rate. Furthermore, the function must be twice differentiable as well in case of Newton method to work, which is not always the case.