**1. (True/False) Stochastic gradient ascent often requires fewer passes over the dataset than batch gradient ascent to achieve a similar log likelihood.**

True

**False**

**True**

False

**2. (True/False) Choosing a large batch size results in less noisy gradients**

**True**

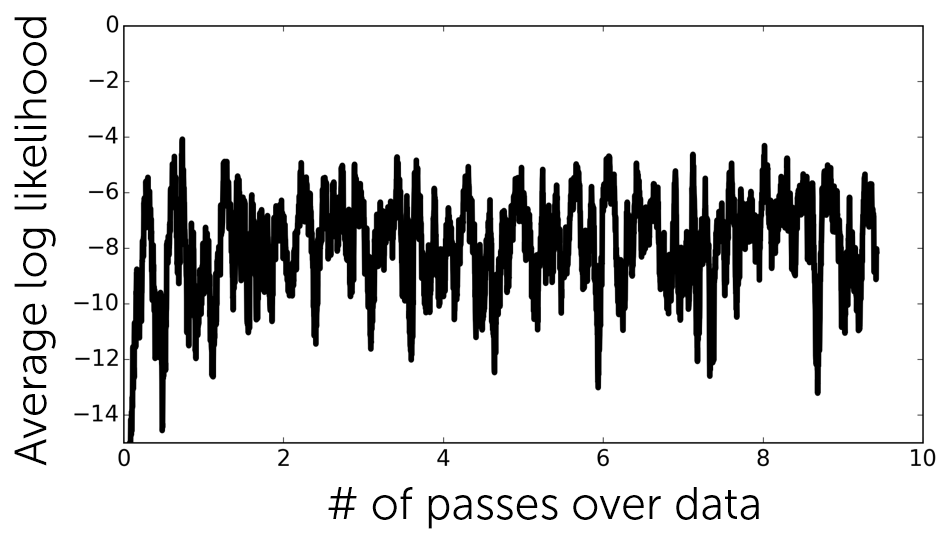
False

**3. (True/False) The set of coefficients obtained at the last iteration represents the best coefficients found so far.**

True

**False**

**4. Suppose you obtained the plot of log likelihood below after running stochastic gradient ascent.**



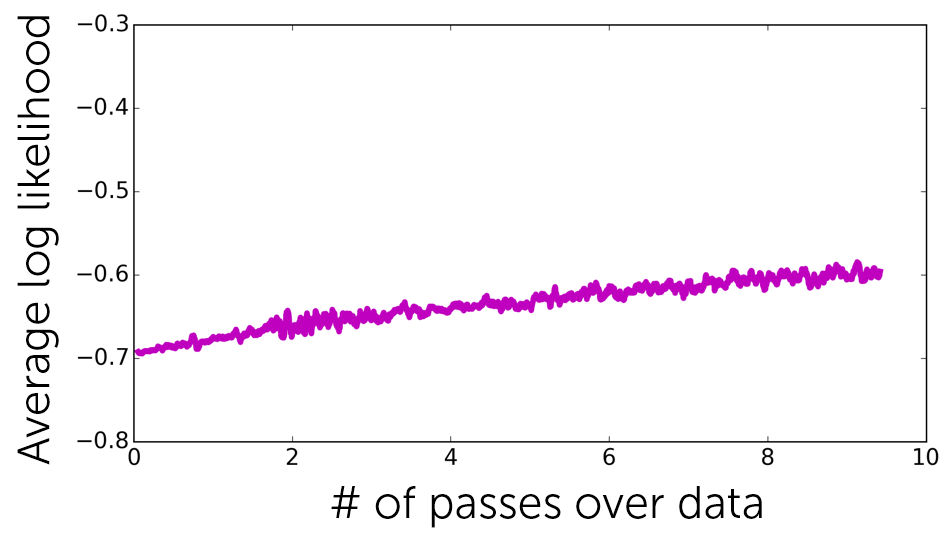
**Which of the following actions would help the most to improve the rate of convergence?**

Increase step size

**Decrease step size**

Decrease batch size

**5. Suppose you obtained the plot of log likelihood below after running stochastic gradient ascent.**



**Which of the following actions would help to improve the rate of convergence?**

Increase batch size

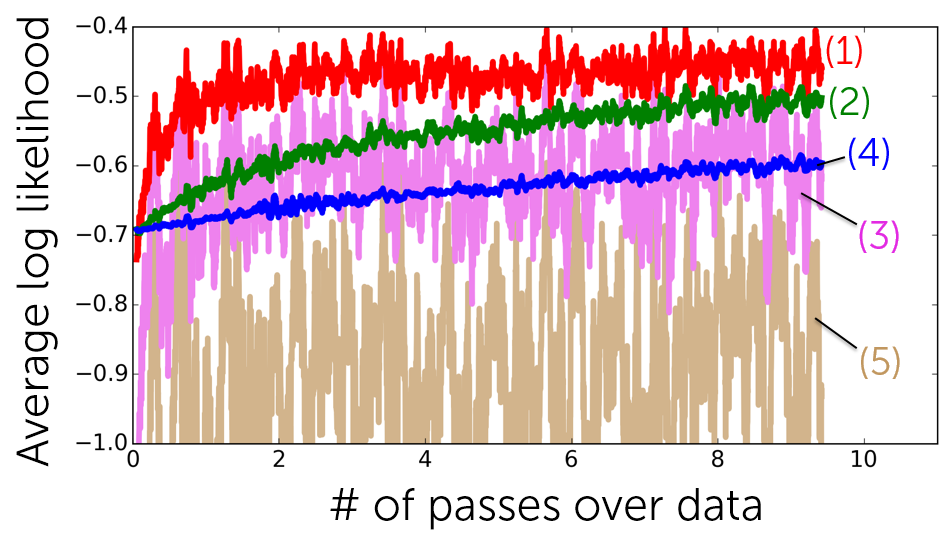
**Increase step size**

Decrease step size

**6. Suppose it takes about 1 milliseconds to compute a gradient for a single example. You run an online advertising company and would like to do online learning via mini-batch stochastic gradient ascent. If you aim to update the coefficients once every 5 minutes, how many examples can you cover in each update? Overhead and other operations take up 2 minutes, so you only have 3 minutes for the coefficient update.**

180000

**7. In search for an optimal step size, you experiment with multiple step sizes and obtain the following convergence plot.**



**Which line corresponds to step sizes that are larger than the best? Select all that apply.**

(1)

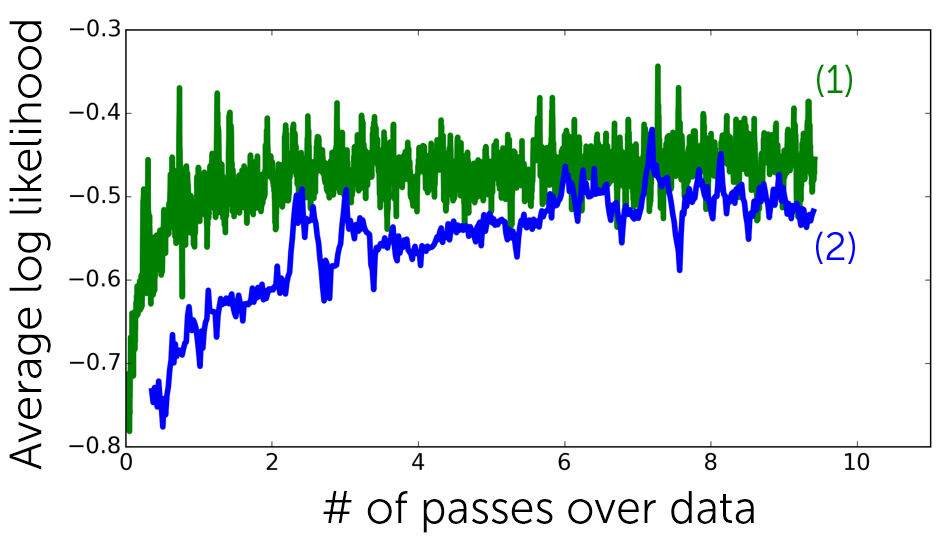
(2)

**(3)**

(4)

**(5)**

**8. Suppose you run stochastic gradient ascent with two different batch sizes. Which of the two lines below corresponds to the smaller batch size (assuming both use the same step size)?**



**(1)**

(2)

**9. Which of the following is NOT a benefit of stochastic gradient ascent over batch gradient ascent? Choose all that apply.**

Each coefficient step is very fast.

**Log likelihood of data improves monotonically.**

Stochastic gradient ascent can be used for online learning.

Stochastic gradient ascent can achieve higher likelihood than batch gradient ascent for the same amount of running time.

**Stochastic gradient ascent is highly robust with respect to parameter choices.**

**10. Suppose we run the stochastic gradient ascent algorithm described in the lecture with batch size of 100. To make 10 passes over a dataset consisting of 15400 examples, how many iterations does it need to run?**

1540