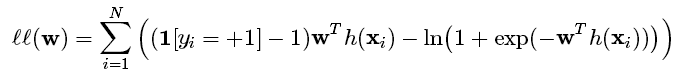
**2. In Module 3 assignment, there were 194 features (an intercept + one feature for each of the 193 important words). In this assignment, we will use stochastic gradient ascent to train the classifier using logistic regression. How does the changing the solver to stochastic gradient ascent affect the number of features?**

Increases

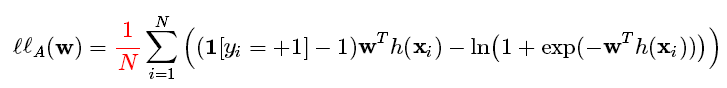
Decreases

**Stays the same**

**3. Recall from the lecture and the earlier assignment, the log likelihood (without the averaging term) is given by**



**whereas the average log likelihood is given by**



**How are the functions ll(w) and ll\_A(w) related?**

ll\_A(w) = ll(w)

**ll\_A(w) = (1/N)\*ll(w)**

ll\_A(w) = N\*ll(w)

ll\_A(w) = ll(w) - ∥w∥

**4. Refer to the sub-section Computing the gradient for a single data point.**

**The code block above computed**



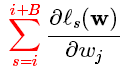
**for j = 1 and i = 10. Is this quantity a scalar or a 194-dimensional vector?**

**A scalar**

A 194-dimensional vector

**5. Refer to the sub-section Modifying the derivative for using a batch of data points.**

**The code block computed**

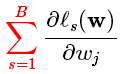


**for j = 10, i = 10, and B = 10. Is this a scalar or a 194-dimensional vector?**

**A scalar**

A 194-dimensional vector

**6. For what value of B is the term**



**the same as the full gradient**



**? A numeric answer is expected for this question. Hint: consider the training set we are using now.**

47780

**7. For what value of batch size B above is the stochastic gradient ascent function logistic\_regression\_SG act as a standard gradient ascent algorithm? A numeric answer is expected for this question. Hint: consider the training set we are using now.**

47780

**8. When you set batch\_size = 1, as each iteration passes, how does the average log likelihood in the batch change?**

Increases

Decreases

**Fluctuates**

**9. When you set batch\_size = len(train\_data), as each iteration passes, how does the average log likelihood in the batch change?**

**Increases**

Decreases

Fluctuates

**10. Suppose that we run stochastic gradient ascent with a batch size of 100. How many gradient updates are performed at the end of two passes over a dataset consisting of 50000 data points?**

1000

**11. Refer to the section Stochastic gradient ascent vs gradient ascent.**

**In the first figure, how many passes does batch gradient ascent need to achieve a similar log likelihood as stochastic gradient ascent?**

It's always better

10 passes

20 passes

**150 passes or more**

**12. Questions 11 and 12 refer to the section Plotting the log likelihood as a function of passes for each step size.**

**Which of the following is the worst step size? Pick the step size that results in the lowest log likelihood in the end.**

1e-2

1e-1

1e0

**1e1**

1e2

1e-2

1e-1

1e0

1e1

**1e2**

**13. Questions 11 and 12 refer to the section Plotting the log likelihood as a function of passes for each step size.**

**Which of the following is the best step size? Pick the step size that results in the highest log likelihood in the end.**

1e-4

1e-2

**1e0**

1e1

1e2