× (?)

Training Logistic Regression via Stochastic Gradient Ascent

Quiz passed!

13/13 points earned (100%)

Back to Week 7 (/learn/ml-classification/home/week/7)

1 / 1 points

1.

Are you using GraphLab Create? Please make sure that

1. You are using version 1.8.3 of GraphLab Create. Verify the version of GraphLab Create by running

`graphlab.version`

inside the notebook. If your GraphLab version is incorrect, see this post (https://www.coursera.org/learn/ml-classification/supplement/LgZ3l/installing-correct-version-of-graphlab-create) to install version 1.8.3. **This assignment is not guaranteed to work with other versions of GraphLab Create.**

2. You are using the IPython notebook named module-10-online-learning-assignment-blank.ipynb obtained from the associated reading.

This question is ungraded. Check one of the three options to confirm.

1 / 1 points

Congratulations!

You have successfully completed **Machine Learning: Classification**, 1 of 6 courses in **Machine Learning** from **University of Washington**.

Final Grade 100.0 %

Unlock a certificate to share your achievement with the world!

Keep Learning

Machine Learning

- **Machine Learning Foundations: A Case Study Approach** 100.0%
- **Machine Learning: Regression** 100.0%
- **Machine Learning: Classification** %
  Case Studies: Analyzing Sentiment & Loan Default Prediction In our case study on analyzing sentiment, you will create models that predict a class...
  View Course Starts April 4 (/learn/ml-classification)
- **Machine Learning: Clustering & Retrieval** --%
- **Machine Learning: Recommender Systems & Dimensionality Reduction** -- %
- **Machine Learning Capstone: An Intelligent Application with Deep Learning** -- %

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?

---

1 / 1
points

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

$$\ell\ell(\mathbf{w}) = \sum_{i=1}^{N} \Big((\mathbf{1}[y_i = +1] - 1)\mathbf{w}^T h(\mathbf{x}_i) - \ln\big(1 + \exp(-\mathbf{w}^T h(\mathbf{x}_i))\big) \Big)$$

whereas the average log likelihood is given by

$$\ell\ell_A(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^{N} \Big((\mathbf{1}[y_i = +1] - 1)\mathbf{w}^T h(\mathbf{x}_i) - \ln\big(1 + \exp(-\mathbf{w}^T h(\mathbf{x}_i))\big) \Big)$$

How are the functions ll(w) and ll_A(w) related?

---

1 / 1
points

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

The code block above computed

$$\frac{\partial \ell_i(\mathbf{w})}{\partial w_j}$$

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

---

1 / 1
points

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

The code block computed

$$\sum_{s=i}^{i+B} \frac{\partial \ell_s(\mathbf{w})}{\partial w_j}$$

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

---

1 / 1
points

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

$$\sum_{s=1}^{B} \frac{\partial \ell_s(\mathbf{w})}{\partial w_j}$$

the same as the full gradient

$$\frac{\partial \ell(\mathbf{w})}{\partial w_j}$$

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

---

1 / 1
points

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.

---

1 / 1
points

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

---

1 / 1
points

9.
When you set batch_size = len(train_data), as each iteration passes, how does the average log likelihood in the batch change?

---

1 / 1
points

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?

---

1 / 1
points

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?

---

1 / 1

points

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.

---

1 / 1
points

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.
