

COMPUTER SUBJECT: TRAINING MODELS

TYPE: GROUP WORK ASSIGNMENTS/DISCUSSION

IDENTIFICATION: CHAPTER 4/MICL

COPYRIGHT: *Michael Claudius*

LEVEL: EASY

DURATION: < 60 min

SIZE: 50 lines!!

OBJECTIVE: Understanding the concepts behind training models

REQUIREMENTS: ML Ch. 4

COMMANDS:

1. Imagine that a function has the complexity $O(m) \times O(n^2)$ where n is number of features and m is number of instances.

Fill out the table for missing values by calculating $m \times n^2$.

m/n	m=100	1.000	10.000	100.000
n=1	100	1000	10.000	100.000
10	10.000	100.000	10^6	
100				
1.000				
10.000				
100.000				

Got it ?

2. Explain Computational complexity and the challenges for the Closed Form Solution
3. Explain the cost function and the purpose of it.

Equation 4-3. MSE cost function for a Linear Regression model

$$\text{MSE}(\mathbf{X}, h_{\theta}) = \frac{1}{m} \sum_{i=1}^m (\theta^T \mathbf{x}^{(i)} - y^{(i)})^2$$

4. Do you understand the normal equation and what it is used for ?

Equation 4-4. Normal Equation

$$\hat{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

In this equation:

- $\hat{\theta}$ is the value of θ that minimizes the cost function.
- \mathbf{y} is the vector of target values containing $y^{(1)}$ to $y^{(m)}$.

5. Which Linear Regression training algorithm can you use if you have a training set with millions of features?
6. Suppose the features in your training set have very different scales. Which algorithms might suffer from this, and how? What can you do about it?
7. Can Gradient Descent get stuck in a local minimum when training a Logistic Regression model?

8. Do all Gradient Descent algorithms lead to the same model, provided you let them run long enough?
9. Suppose you use Batch Gradient Descent and you plot the validation error at every epoch. If you notice that the validation error consistently goes up, what is likely going on? How can you fix this?
10. Is it a good idea to stop Mini-batch Gradient Descent immediately when the validation error goes up?
11. Which Gradient Descent algorithm (among those we discussed) will reach the vicinity of the optimal solution the fastest? Which will actually converge? How can you make the others converge as well?
12. Suppose you are using Polynomial Regression. You plot the learning curves and you notice that there is a large gap between the training error and the validation error. What is happening? What are three ways to solve this?
13. Suppose you are using Ridge Regression and you notice that the training error and the validation error are almost equal and fairly high. Would you say that the model suffers from high bias or high variance? Should you increase the regularization hyperparameter α or reduce it?
14. Why would you want to use:
 - a. Ridge Regression instead of plain Linear Regression (i.e., without any regularization)?
 - b. Lasso instead of Ridge Regression?
 - c. Elastic Net instead of Lasso?
15. Suppose you want to classify pictures as outdoor/indoor and daytime/nighttime. Should you implement two Logistic Regression classifiers or one Softmax Regression classifier?
16. Implement Batch Gradient Descent with early stopping for Softmax Regression (without using Scikit-Learn).