**[Machine Learning] hw4** B01902040

**鍾毓安**

1. (b)

We know that target complexity acts like noise. If we choose to use a less powerful learning model to fit the training data, in general, the result (at least ) won’t be that good as a powerful one. In this problem, , which means that is less powerful than , so the deterministic noise will increase.



1. (c)



1. (a), (d)
2. We compute the gradient of the augmented error and update the weight by gradient descent.



(d) Just another form of (a) :



1. (b), (c)

As mentioned in lecture (Lecture 14, page 11), when is getting larger, the model prefers shorter (to keep the penalty small).



1. (c)

Given 3 data points : .



For hypothesis :



Leaving out, the training result is . Therefore, .



Leaving out, the training result is . Therefore, .



Leaving out, the training result is . Therefore,



For hypothesis :



Leaving out, the training result is . Therefore, .



Leaving out, the training result is . Therefore, .



Leaving out, the training result is . Therefore,



Now we have to solve for .



Since must , ,



and thus , but note that , .



1. (a), (c)
2. There are win-lose predictions for games.



1. Since there are 32 win-lose predictions, the sender should target to begin with at least 32 people.
2. After knowing the result of the first game, the sender could cross out half of people (16 people left), who just received the wrong predictions for the first game.
3. The sender should target on 32 people on the first week, and every week after the game, he could decrease his targets by half, that is, 16 people on the second week, 8 people on the third … Therefore, at least letters should be sent.



1. (b)

Besides the 62 letters described in 6.(d), 1 more letter should be sent to ask the target person, who received correct predictions on all 5 games, whether he wish to see next week’s prediction or not, letters are required.



.



1. (a)

, since I did mathematical derivations and came up with a credit approval function just once without peeping any data.

1. (c)



1. (a), (c)

The key concept in this problem is that the original distribution of customers has been ‘modified’, since (1) the bank applied vague idea  to determine whether to give a customer credit card or not, and (2) only those who had been given credit cards were monitored.

The distribution looks like this :

Original Customer Distribution



Approved

(We have data for this part)

Disapproved

1. If the old credit approval function was (approve all customers), then there’s no branches existed since all customers were directed to the “ Approved ” part.
2. Only by applying  AND  together, the procedure of approving

credit to customers would be directed to the right scenario where I came up with .

1. (d)

To obtain the optimal , just simply set and solve this linear equation. Therefore, we derive .



1. (b)

In the lecture, we have derived the optimal solution for . Therefore, we now have :



When , turns out to be same as , while other options aren’t always the cases.



**[Bonus]**

21.



From problem 11, we can thus have , so we can then choose  as virtual examples.

22.



From problem 11, we can thus have , so we can then choose  as virtual examples.