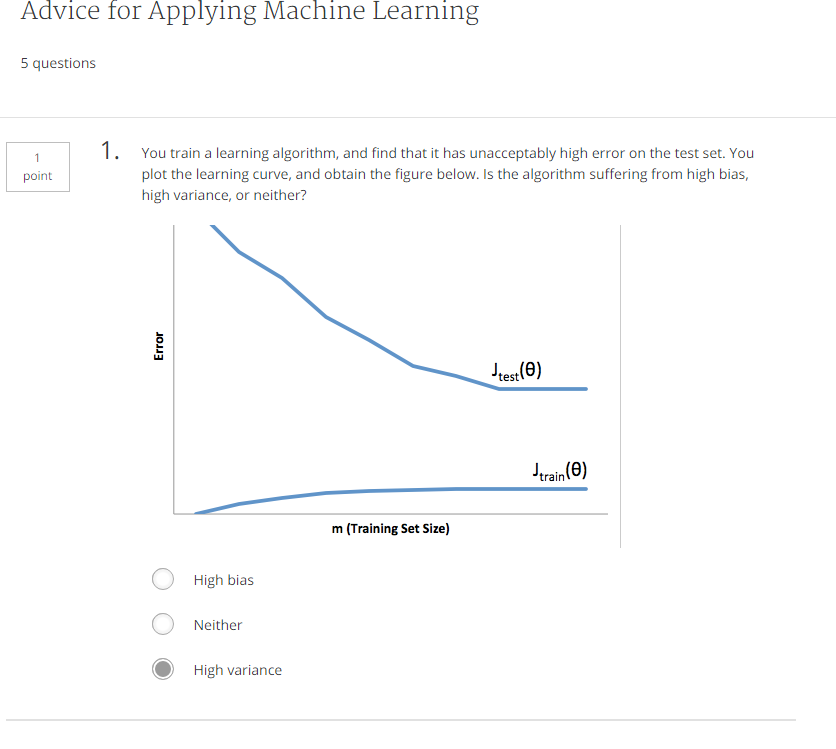
1) one way to evaluate result  is using miss classified error

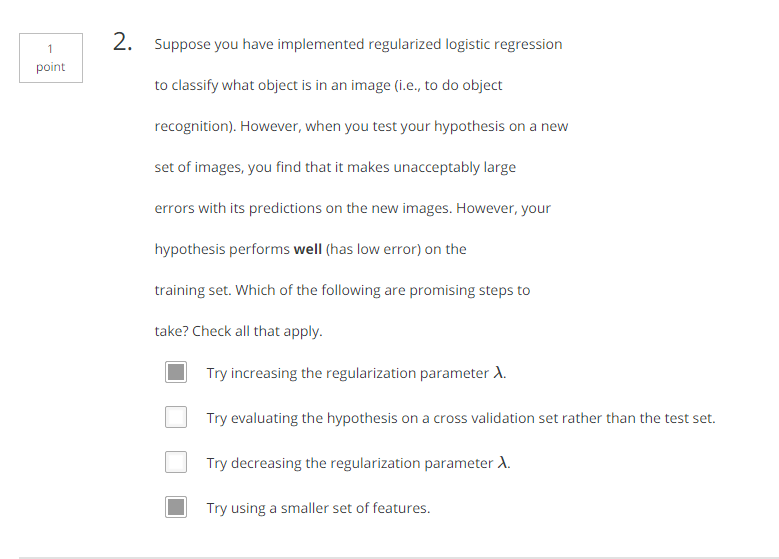
2) Divided data set into 60% training samples, 20% cross validation, 20% testing samples. To check training samples, lambda, number of hidden unit..., we can follow these step Training in training samples, to get theta. Applying to CV, choose the best theta which responding to the lowest cost function. Evaluate in test set.

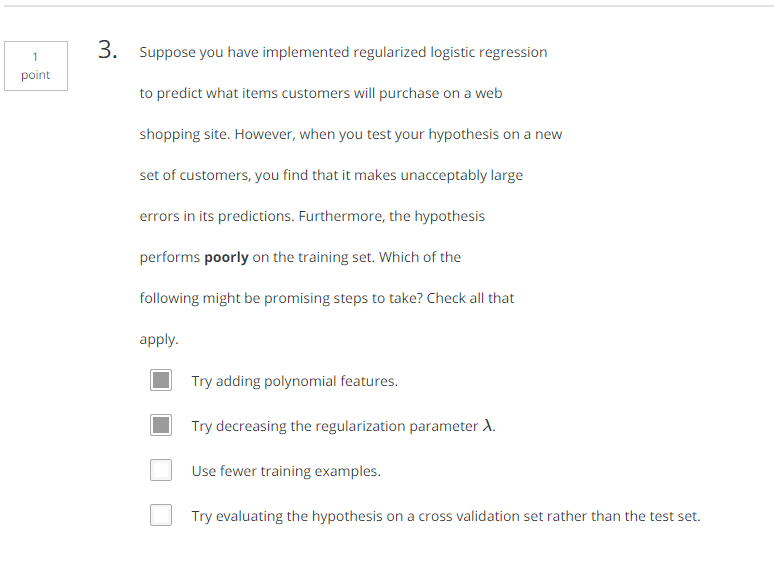
3) bias/ and variance: when higher the d (degree of polynomial), the Jtest is smaller, Jcv smaller into particular point (reduce under fitting problem), then higher again (this time is over-fitting).

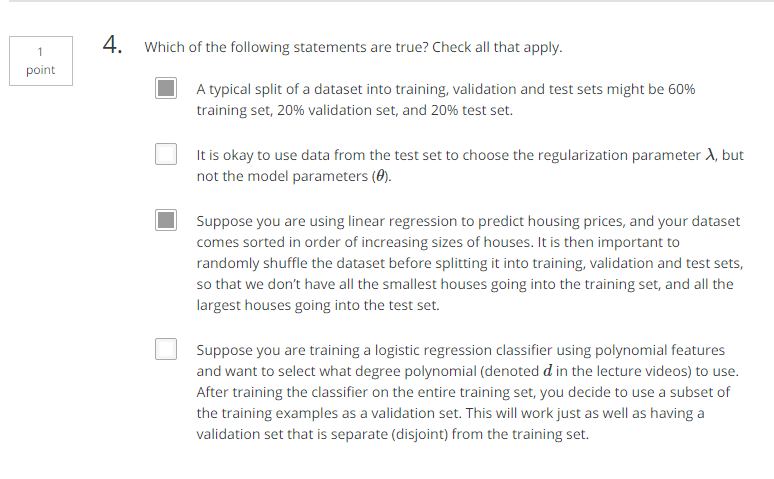
4) Intuition:  more high variant (more over fitting) -> more sensitive with data -> add more data is good.  In the other hand, more high bias (under fitting)-> more rigid-> adding more data does not help

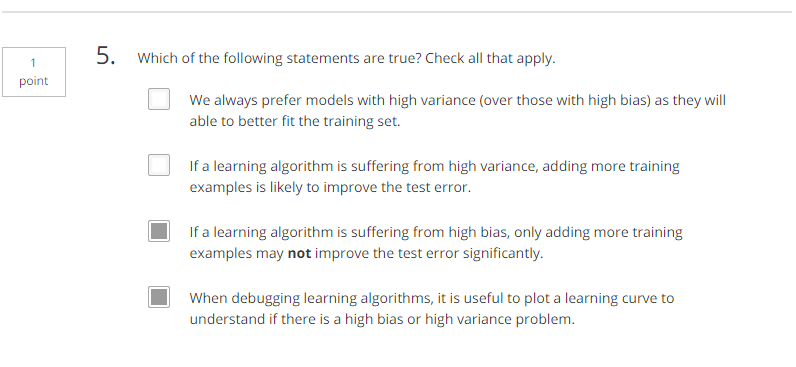
5) Due with the over-fitting problem (high variance), add more training samples or reduce features is difficult. The easier way is adjusting the lambda (regulization part)

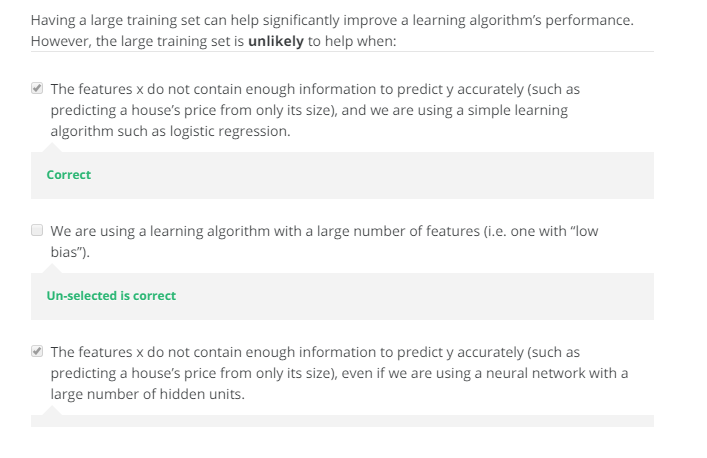




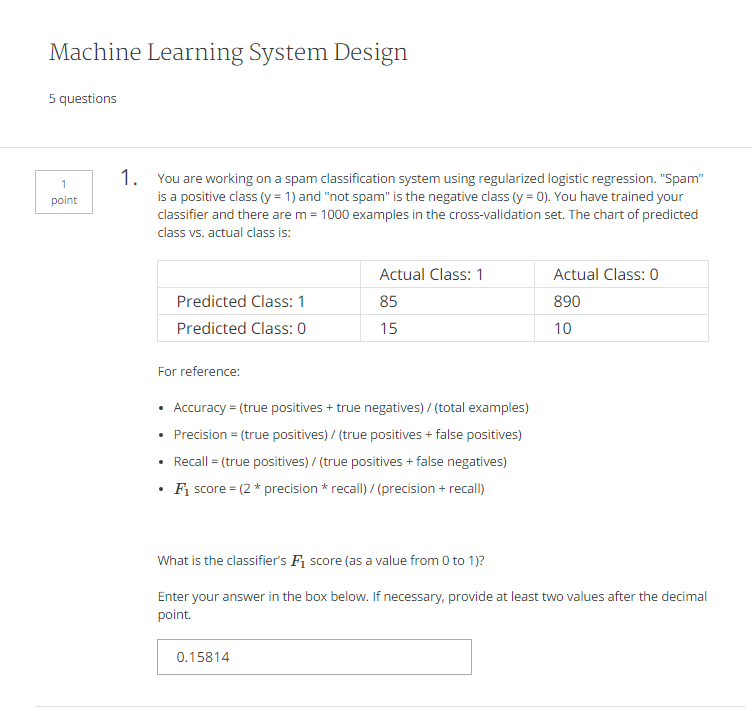


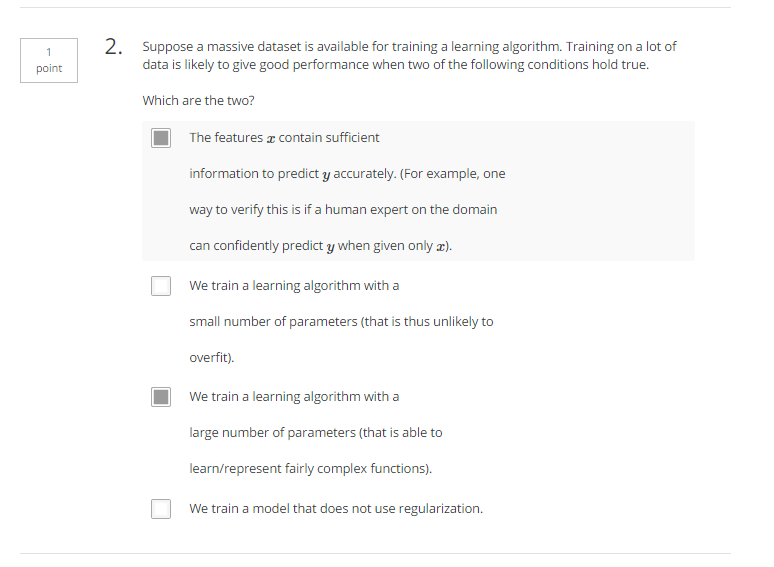


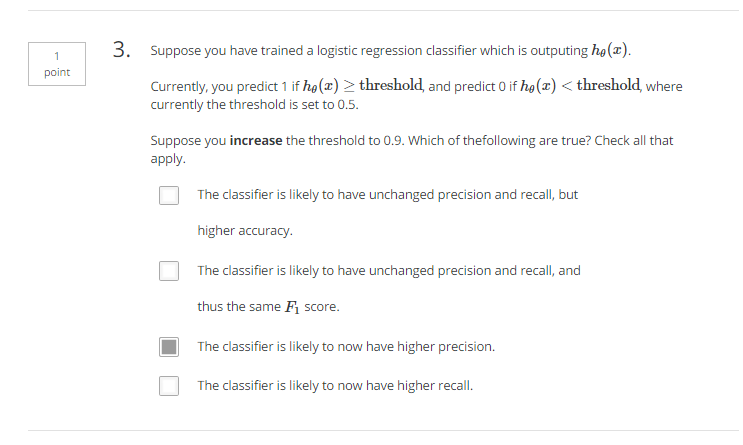


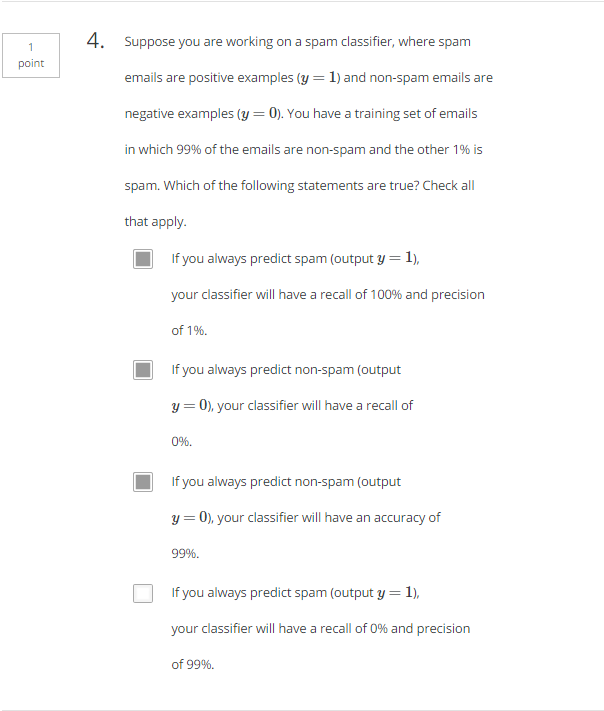


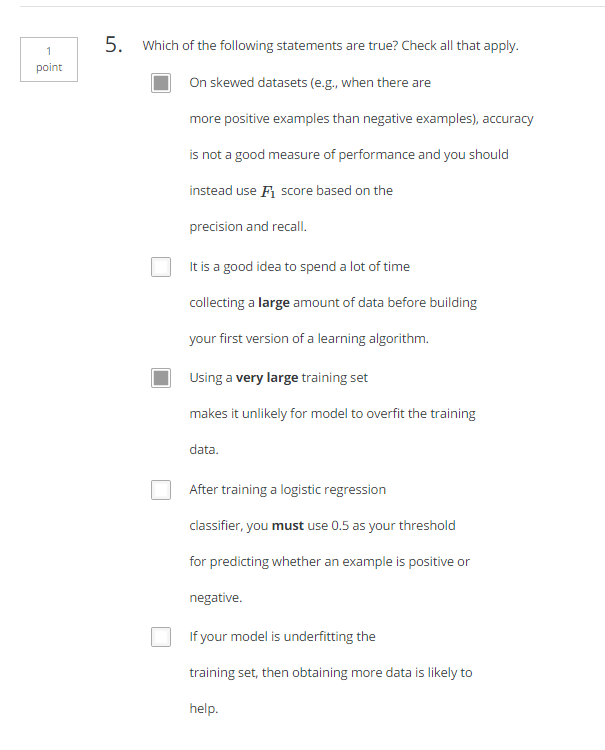
Quiz part 2: câu 4 khó











Giải thích ý 3 câu 5: dung data set lớn thì sẽ làm model training ko bị overfit nữa

Giải câu 4:

Positive example: y=1 spam

Negative example: y=0 non-spam

Dataset: 1%: spam(1)

99%: Nonspma (0)

Th1: y=0

|  |  |  |  |
| --- | --- | --- | --- |
| Actual | | | |
| Predicted |  | 1 | 0 |
| 1 | 0(actual 1, predict 1) | 0 (false positive) |
| 0 | 1(false negative) | 99 |

P= 0/(0+0); ko co

Recall = : 0/(0+1)=0%

Accuracy: 99 nonspam là đúng, (tổng đường chéo- dự đóan đúng 0+99)/ tong so data =100=99%. 1 cai spam ma du doanla nonspam la sai

Th2 y=1:

|  |  |  |  |
| --- | --- | --- | --- |
| Actual | | | |
| Predicted |  | 1 | 0 |
| 1 | 1 | 99 |
| 0 | 0 | 0 |

P=1%

R=100%

Ghi chú:

Precision cao: đã dự đoán là dự đoán đúng

Recall: có nhiều người bị, nhưng chỉ đoán ít người (do chắc ăn mới đoán).  
=> dự đoán ít nhưng đã dự đoán là đúng

Precision thấp: dự đoán mà tỉ lệ đúng ko cao

Recall: có nhiều người bị, đều dự đoán dc gần hết (do có khả năng là cho bị luôn)

* Thà dự đoán sai còn hơn bỏ sót.