X Try again once you are ready.

Required to pass: 80% or higher You can retake this quiz up to 3 times every 8 hours.





You are working on a spam classification system using regularized logistic regression. "Spam" is a positive class (y = 1) and "not spam" is the negative class (y = 0). You have trained your classifier and three are m = 1000 examples in the cross-validation set. The chart of predicted class vs. actual class is:

	Actual Class: 1	Actual Class: 0
Predicted Class: 1	85	890
Predicted Class: 0	15	10

For reference:

- Accuracy = (true positives + true negatives) / (total examples)
- Precision = (true positives) / (true positives + false positives)
- F₁ score = (2 * precision * recall) / (precision + recall)

What is the classifier's accuracy (as a value from 0 to 1)?

Enter your answer in the box below. If necessary, provide at least two values after the decimal point.

0.15814



Suppose a massive dataset is available for training a learning algorithm. Training on a lot of data is likely to give good performance when two of the following conditions hold true.



Which are the two?

We train a model that does not use regularization.

Un-selected is correct

We train a learning algorithm with a

small number of parameters (that is thus unlikely to

Un-selected is correct

We train a learning algorithm with a

large number of parameters (that is able to

learn/represent fairly complex functions).

Correct ...
You should use a "low bias" algorithm with many parameters, as it will be able to make use of the large dataset provided. If the model has too few parameters, it will underfit the large training set.

lacksquare The features x contain sufficient

information to predict \boldsymbol{y} accurately. (For example, one

way to verify this is if a human expert on the domain

can confidently predict y when given only x).

CorrectIt is important that the features contain sufficient information, as otherwise no amount of data can solve a learning problem in which the features do not contain enough information to make an accurate prediction.



Suppose you have trained a logistic regression classifier which is outputing $h_{ heta}(x)$.

Currently, you predict 1 if $h_{\theta}(x) \geq$ threshold, and predict 0 if $h_{\theta}(x)lt$ threshold, where currently the threshold is set to 0.5. Suppose you **decrease** the threshold to 0.1. Which of the following are true? Check all that apply.

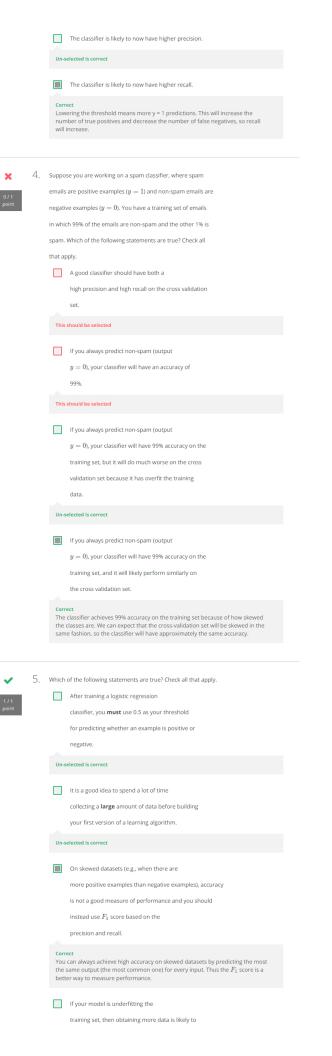
The classifier is likely to have unchanged precision and recall, but

lower accuracy.

The classifier is likely to have unchanged precision and recall, but

higher accuracy.

Un-selected is correct



help.

Using a **very large** training set

makes it unlikely for model to overfit the training

data.

Correct
A sufficiently large training set will not be overfit, as the model cannot overfit some of the examples without doing poorly on the others.

6 8 P