

## Congratulations! You passed!

TO PASS 80% or higher



GRADE 100%

## **Machine Learning System Design**

LATEST SUBMISSION GRADE 100%

1. 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 there are m = 1000 examples in the cross-validation set. The chart of predicted class vs. actual class

1 / 1 point

	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)
- Recall = (true positives) / (true positives + false negatives)
- $F_1$  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.

0.095

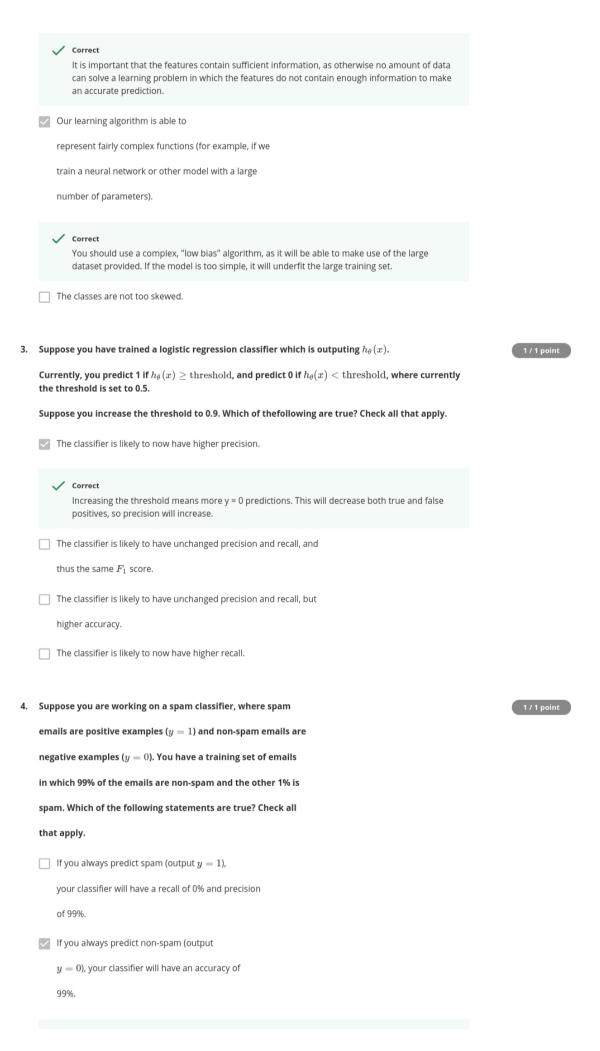


The classifier correctly predicted the true positives and the true negatives = 85 + 10, so the accuracy is 95/1000 = 0.095

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?

	When we are willing to include high
	order polynomial features of $x$ (such as $x_1^2$ , $x_2^2$ ,
	$x_1x_2$ , etc.).
<b>/</b>	A human expert on the application domain
	can confidently predict $\boldsymbol{y}$ when given only the features $\boldsymbol{x}$
	(or more generally, if we have some way to be confident
	that $\boldsymbol{x}$ contains sufficient information to predict $\boldsymbol{y}$
	accurately).



		Correct Since 99% of the examples are y = 0, always predicting 0 gives an accuracy of 99%. Note, however, that this is not a good spam system, as you will never catch any spam.
	<b>~</b>	If you always predict spam (output $y=1$ ), your classifier will have a recall of 100% and precision of 1%.
		Correct Since every prediction is y = 1, there are no false negatives, so recall is 100%. Furthermore, the precision will be the fraction of examples with are positive, which is 1%.
	<b>Y</b>	If you always predict non-spam (output $y=0$ ), your classifier will have a recall of $0\%.$
		Correct Since every prediction is y = 0, there will be no true positives, so recall is 0%.
5.	Wh	ich of the following statements are true? Check all that apply.
		If your model is underfitting the training set, then obtaining more data is likely to
		help.  It is a good idea to spend a lot of time
		your first version of a learning algorithm.
	<b>Y</b>	Using a <b>very large</b> training set  makes it unlikely for model to overfit the training  data.
		<ul> <li>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.     </li> </ul>
		The "error analysis" process of manually examining the examples which your algorithm got wrong can help suggest what are good steps to take (e.g., developing new features) to improve your algorithm's performance.
		Correct This process of error analysis is crucial in developing high performance learning systems, as the space of possible improvements to your system is very large, and it gives you direction about what to work on next.
		After training a logistic regression  classifier, you <b>must</b> use 0.5 as your threshold  for predicting whether an example is positive or

negative.