Comments on Exercise 1

Theoretical Assignments.

1. In parts a and b, the fact that the first derivative is 0 is not enough to show a maximum. It should be verified using the second derivative. Maximizing (and taking the derivative of) the log of an expression is the same as maximizing the expression, and in this case, easier. In part c, it is required to find the distribution of lambda, not a single maximum a posteriori value of lambda. Some people missed the fact that there are m sample points, not just one. Many people worked hard to show that the constant of the distribution of lambda is the same as the constant of the Gamma distribution. This is not needed, because if the distribution is *proportional* to a Gamma distribution, as it clearly is, they have to be equal.
2. Some people thought that H is finite. Some gave a proof that for any single h, the error on the sample is close to the real error with high probability, instead of proving that this holds for **all hypotheses simultaneously**. There were relatively many inaccuracies and arithmetic mistakes.
3. Many answers contained too much hand waving to enable checking. The fact that the x's are not necessarily distinct was noticed only by few.

Programming Assignment.

1. Some submissions were very slow. A properly written program in Matlab should take less than a second to run, even when going over all possible training set sizes.
2. Many people started with a relatively large training set fraction (e.g. 0.2), so their plots did not show the major improvement in the learning curve.
3. Many submissions multiplied the gaussian probabilities of the 4 features and the prior. This is bad practice, because if there are many features, this may lead to numeric issues. The calculation should be done by adding logs of probabilities.
4. Some people had a bug with small training sets, where one or more classes is not present in the training set.
5. Many people relied on Matlab and R to handle calculations involving division by zero and log of zero properly for them. This is dangerous coding practice, and should be done carefully and commented carefully, and best not done at all.
6. It should be noted that there are two legitimate ways to compute the standard deviation of N samples (dividing by N and dividing by N-1) and that either is OK. This choice is sometimes made implicitly through the use of normpdf/dnorm, and one should be aware which choice is being made.

Important: other points are covered in the submission guidelines.