MML Final Review

The Final exam will be during finals week on Monday, May 5 at 11:30 AM. The objective is just going to be to have a few conceptual questions to tie some things together. Here are a few examples.

1. For each of the following pairs of concepts, state the difference between the two items and give and example of each.

   - Regression vs classification
   - Supervised vs unsupervised algorithms
   - Parametric vs Nonparametric algorithms

2. Recall that a *perceptron* is a neural network with no hidden layers. Let's suppose that we have a perceptron with two x inputs, a constant node, and one y output.

   a. Draw the corresponding network. Be sure to label all the nodes appropriately, including subscripts as appropriate. Label the edges as appropriate as well.
   b. Assuming a generic activation function φ, indicate how your labels combine to form value at y.
   c. Indicate how you might choose the activation function to show that the perceptron includes
      i. Linear regression and
      ii. Logistic regression.

3. If I had to name the top three applications of linear algebra to machine learning, I might just say

   - i. Norms,
   - ii. Orthogonal projection, and
   - iii. Diagonalization of a matrix A to SDS^{-1} using eigenvalues and eigenvectors.

Here's one question related to each of those:

- Which norm used to measure error in a least square approximation and how?
- How is orthogonal projection related to the normal equations and the solution of least squares problems?
- How is diagonalization used in Principal Component Analysis?

4. Suppose we'd like to build a machine learning algorithm for each of the following situations:

   a. A medical diagnosis system to use patient similarity compared to past cases (based on features like age, symptoms, and blood pressure). The objective is to predict whether a patient has a certain condition.
   b. An online advertising platform to estimate the probability that a user will click on an ad, based on features like time of day, device type, and ad category.
   c. A model to predict the number of points that this team might defeat that team by based on prior performances.
   d. An algorithm for Spotify to predict song popularity. The measure of popularity might be number of streams per week and the predictors might be a number of disparate properties like tempo, genre, duration, release season, artist popularity, etc.

   Identify which of the following techniques might work best for each of these:

   - Linear regression,
   - Logistic regression,
   - KNN Classifcation, or
   - KNN Regression.

   Of course, we'd like to know why.

5. Often, we use the notation of linear algebra to write formulae much more compactly. In linear regression, for example, we derive a formula of the form

$$\left(\sum_{i=1}^{p} a_i x_i \right) + b.$$

   We might write that, though, as
$$X^T A + b,$$

   where X is the vector of variables and A is the vector of parameters.

   Now, consider the neural network shown in figure 1. Let's suppose that layer i has the activation function g_i. How might you express the value coming out of layer i as a function of the inputs from layer $i - 1$

   a. Using summation notation? And
   b. Using more compact matrix vector notation?

6. The images shown in figure 2 display the results of two different classification algorithms applied to the same data set. The data set consists of 300 different two-dimensional data points each indicated by color. What kinds of algorithms might have generated those plots? There might be more than one reasonable response for each plot but you should be able to articulate some reason for your response.
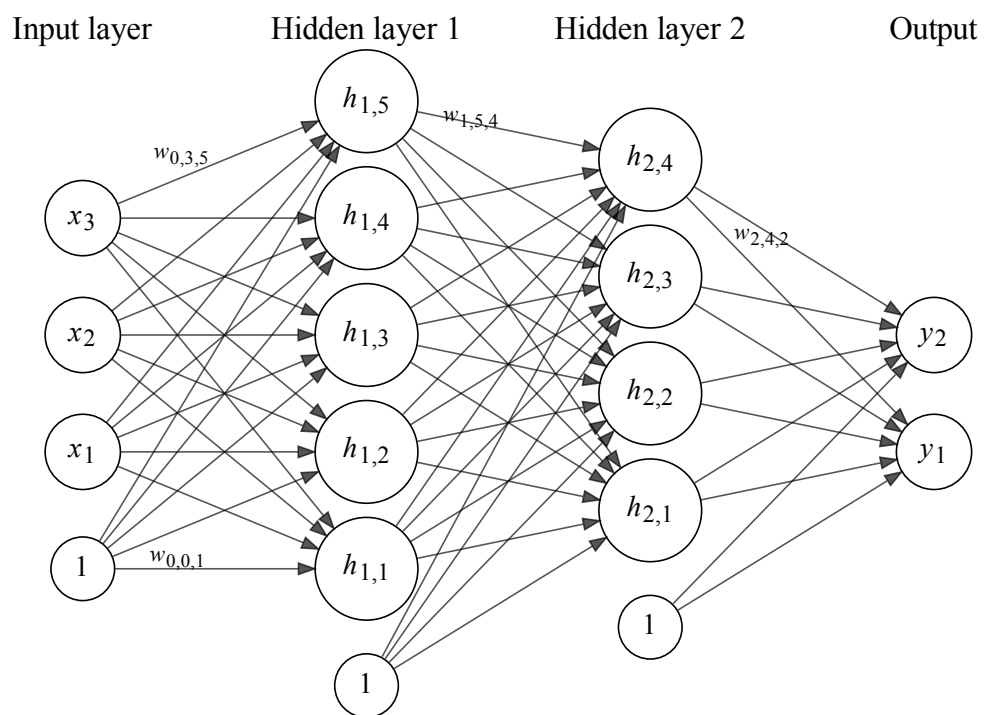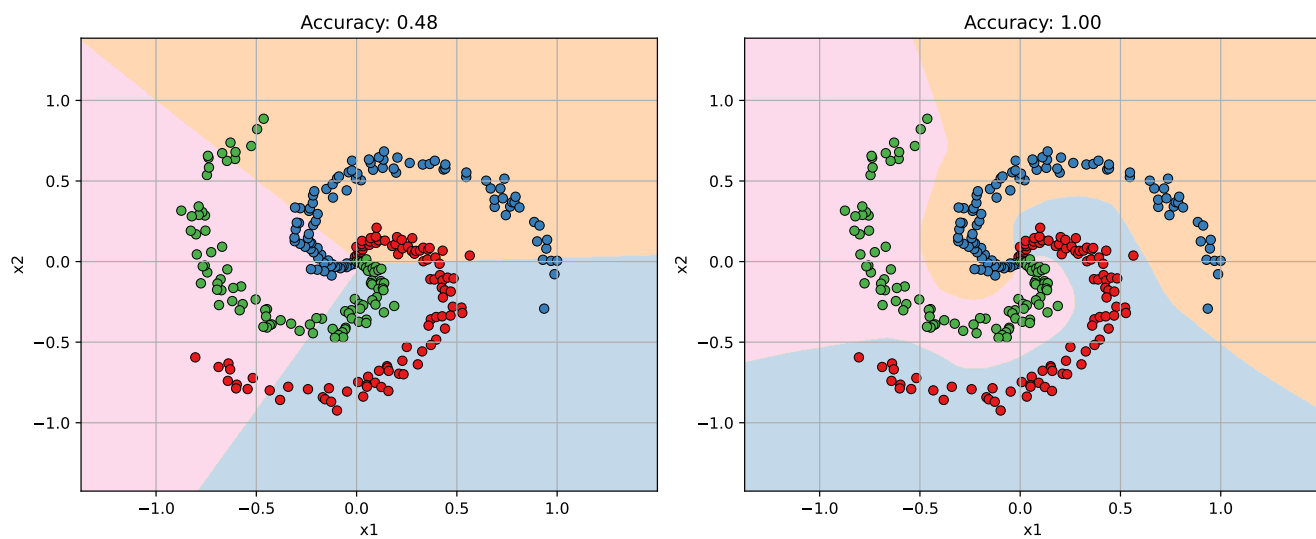
Figures

Figure 1: A labeled neural network

Figure 2: Classifcation plots