1. (10 points total) Mixture of experts.

    (a) (2 points) Draw the mixture of experts architecture.

    (b) (5 points) Explain the architecture.

(c) (3 points) Explain the simple error function used for training, $E = \sum_i p_i(d - y_i)^2$

2. (14 points total) Clustering

   (a) (2 points) Describe the two steps that the k-means algorithm alternates between.

   (b) (1 point) Give an example in which the k-means algorithm finds a bad local optimum.

(c) (2 points) Explain the mixture-of-Gaussians generative model; that is, describe how to generate data for the model.

(d) (4 points) In the EM algorithm for fitting a mixture of Gaussians to data, what does the E-step do? What does the M-step do? Use as little mathematics in your explanation as possible.

(e) (3 points) In what ways does the EM algorithm improve upon k-means?

(f) (2 points) Write down the formula for μ_i^{new}, the updated mean of Gaussian i in the EM algorithm. Give a simple, intuitive explanation of the formula.

3. (10 points total) Recurrent Networks.

   Consider the following recurrent neural network, in which the bottom node is the input, the top node is the output, and the middle two nodes are hidden.
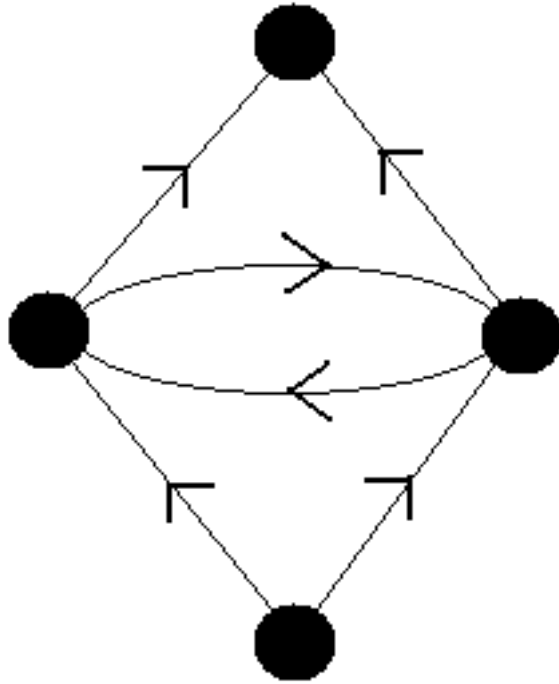
   Draw an equivalent, non-recurrent neural network representing three time instances. Let x_i and y_i be the input and output, respectively, to the recurrent network at time i. In your equivalent, non-recurrent network, label the input and output nodes with x_i and y_i for appropriate values of i.

This page is for answers and rough work

4. (16 points total) Restricted Boltzman Machines

    (a) (3 points) What is a restricted Boltzman machine (RBM)? (You may define it in terms of general Boltzman machines.) Draw a picture of simple RBM.

    (b) (2 points) What properties does the probability distribution of a RBM have that a general Boltzman machine does not?

(c) (4 points) How do these properties affect the Markov Chain Monte Carlo procedure for reaching thermal equilibrium of the hidden and visible units?

(d) (2 points) How do these properties affect the Markov Chain Monte Carlo procedure for reaching thermal equilibrium of the hidden units given values for the visible units?

(e) (5 points) Describe a procedure for learning the weights of a restricted Boltzman machine.

5. (10 points total) Stacking Restricted Boltzman Machines

    (a) (6 points) Let RBM1, RBM2 and RBM3 be three restricted Boltzman machines, and suppose we stack them on top of each other, with RBM1 on the bottom, RBM2 in the middle, and RBM3 on top. Describe a procedure for learning the weights of the stacked RBMs.

(b) (4 points) Describe how to generate data with the stacked RBMs.

6. (12 points total) Autoencoders

   (a) (3 points) Describe the structure of an autoencoder, and draw a simple picture of one.

   (b) (1 point) During learning by backpropagation, what is the target output of an autoencoder?

   (c) (3 points) How does an autoencoder achieve dimensionality reduction?

(d) (5 points) Describe how to use restricted Boltzman machines to pretrain a deep autoencoder. How does learning proceed after pretraining? Feel free to use pictures in your explanation.

7. (15 points total) Support Vector Machines

    (a) (2 points) What does it mean for a data set to be linearly separable?

    (b) (1 point) What is the margin of a linear separator?

    (c) (2 points) Why is it important to maximize the margin?

(d) (5 points) Give a simple example of how linear classification in a high-dimensional space can give rise to non-linear classification in a low-dimensional space.

(e) (5 points) What is the kernel trick, and how does it allow an SVM to do linear classifi-
cation in very high-dimensional feature spaces?

This page is for answers and rough work.

This page is for answers and rough work.