Name ____________________________________

CSE 5526 - Autumn 2014

Practice Final

Directions for the real final exam on 12/12, 10:00am-11:45am:

1. The real final will be a closed-book, closed-note test, except for a single 8.5x11" sheet of notes that you are allowed to bring.
2. The final is comprehensive, but focused on material covered since the midterm
3. You may not consult with any other person.
4. You may not use any communication device or computer
5. You have 105 minutes to finish.
6. Write all work on the test paper. Use reverse side if needed (clearly indicate so).
7. There are *four* problems, with a total of 100 points, *plus* a bonus problem (10 points)

Problem 1. Short Answers (25 points)

(a) (4 points) What is the problem with a learning rate that is too large? Too small? What would be a good way to set the learning rate adaptively?

(b) (4 points) What are the two terms of the Boltzmann machine update each trying to accomplish?

(c) (4 points) Why is the maximum margin hyperplane better than other separating hyperplanes?

(d) (4 points) How are SVMs similar to RBF networks? How are they better?

(**e**) (4 points) If an MLP were going to predict information about a person from a set of features computed on their credit history, how many and what type of output units should it use and what error function should it minimize to predict a numerical score from 0 to 800? What about to predict what city they were born in?

(**f**) (4 points) What is the restriction in a restricted Boltzmann machine? Why are restricted Boltzmann machines easier to train than general Boltzmann machines?

Problem 2. (25 points) Consider a Boltzmann machine with three neurons, labeled, A, B, and C.

(a) Write down the energy of all 8 configurations of the network if the weights are set to

$$w_{AB} = \ln 2\,, w_{AC} = \ln 4\,, w_{BC} = \ln 8$$

A	**B**	**C**	$\boldsymbol{E(A,B,C)}$	$\boldsymbol{\tilde{p}(A,B,C)}$	$\boldsymbol{p(A,B,C)}$
1	1	1			
1	1	-1			
1	-1	1			
1	-1	-1			
-1	1	1			
-1	1	-1			
-1	-1	1			
-1	-1	-1			

(b) Write the corresponding unnormalized probabilities, $\tilde{\boldsymbol{p}}(A, B, C)$, in the appropriate column above for each state at temperature 1.

(c) Compute the partition function $\boldsymbol{Z}$ for temperature 1.

(d) Write the corresponding normalized probabilities, $p(A, B, C)$, in the appropriate column above for each state at temperature 1.

(e) What happens to the probabilities at temperature 0? What happens at temperature ∞?

Problem 3. (25 points) Suppose we train a support vector machine and it identifies N support vectors, each of which is represented by a K-dimensional feature vector, including one element that is always 1. We then test this SVM on M data points.

(a) If the SVM was trained using a linear kernel, write down the function that this SVM uses to classify a new data point

(b) If the SVM was trained using a radial basis function kernel, write down the function that this SVM uses to classify a new data point

(c) How many evaluations of the kernel function are necessary to classify M test points for the RBF kernel?

(d) How many evaluations of the kernel function are necessary to classify M test points for the linear kernel? Can it be done in fewer than the RBF?

Problem 4. (25 points) Consider the function $f(\mathrm{x,y}) = -\frac{5}{2}x^2 + 3xy - \frac{5}{2}y^2 - 2x - 2y$ and the constraint $g(\mathrm{x,y}) = y$

(a) Write down the Lagrangian function, $L(\mathrm{x,y,\lambda})$, for maximizing $f(\mathrm{x,y})$ subject to $g(\mathrm{x,y}) = 0$ with lagrange multiplier λ

(b) Write down the partial derivatives of $L(\mathrm{x,y,\lambda})$ with respect to x, y, and λ

(c) Solve this set of three equations for the three unknowns x, y, and λ

(d) Find the values of x, y, and λ that maximize $f(x,y)$ subject to $g(\mathrm{x,y}) \geq 0$

(e) Find the values of x, y, and λ that maximize $f(x,y)$ subject to $g(\mathrm{x,y}) \leq 0$

Bonus Problem. (10 points) Show that Mercer kernels satisfy the Cauchy-Schwarz inequality

$$k(\boldsymbol{x}, \boldsymbol{x}')k(\boldsymbol{x}', \boldsymbol{x}) \leq k(\boldsymbol{x}, \boldsymbol{x})k(\boldsymbol{x}', \boldsymbol{x}')$$

Hint: use the determinant of a 2×2 Gram matrix