Name _____

CSE 5526 - Autumn 2014

# Final Exam

**Dec 12, 2014, 10:00am-11:45am**

**Directions:**

1. This is a closed-book, closed-note exam, except for a single 8.5x11" sheet of notes.
2. It 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)** (5 points) Are the following models used for supervised or unsupervised learning (put an X in the relevant column for each algorithm)?

| **Algorithm** | **Supervised** | **Unsupervised** |
|---|---|---|
| Radial basis function network | | |
| McCulloch-Pitts neuron | | |
| Support vector machine | | |
| Deep neural network | | |
| Self-organizing map | | |
| Restricted Boltzmann machine | | |
| Multi-layer perceptron | | |
| Boltzmann machine | | |
| Linear regression | | |
| Deep belief network | | |

**(b)** (5 points) What is the relationship between the maximum margin hyperplane and a support vector?

**(c)** (5 points) How are SVMs similar to M-P neurons trained using the perceptron rule? How are they better?

**(d)** (10 points) Please provide descriptive answers (no formulas necessary) to the following 4 questions for **both** SVMs using Gaussian (RBF) kernels **and** RBF networks.

1) How do we choose the location of the Gaussian centers?

2) How do we determine the widths of the Gaussians?

3) How do we determine the weights $w_j$ on the output of each basis function?

4) How do we select the number of Gaussian centers?

**Problem 2.** (25 points) Suppose we train a support vector machine on N data points, each of which is represented by a K-dimensional feature vector, including one element that is always 1.

**(a)** (7 points) Write down the primal, constrained (not Lagrangian) optimization for this SVM. How many dimensions is this quadratic program optimizing?

**(b)** (6 points) Write down the dual Lagrangian optimization for this SVM, including constraints. How many dimensions is this quadratic program optimizing?

**(c)** (6 points) If solving a quadratic program takes time proportional to the cube of the number of variables (and is independent of the number of constraints), approximately how long would it take to train an SVM for a spam classification task where K=100 and N=1e7 using the primal formulation? Using the dual formulation?

**(d)** (6 points) Approximately how long would it take to train an SVM for an fMRI classification task where K=1e5, N=1000 using the primal formulation? Using the dual formulation?

**Problem 3.** (25 points) Consider the function $f(\mathrm{x},\mathrm{y}) = -9x^2 - y^2 + 18x - 2y$ and the constraint $g(\mathrm{x},\mathrm{y}) = y - x$

**(a)** (5 points) Write down the Lagrangian function, $L(\mathrm{x},\mathrm{y},\lambda)$, for maximizing $f(\mathrm{x},\mathrm{y})$ subject to $g(\mathrm{x},\mathrm{y}) = 0$ with lagrange multiplier $\lambda$

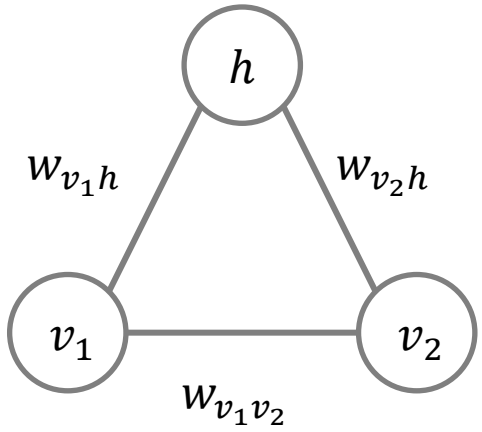
**(b)** (5 points) Write down the partial derivatives of $L(\mathrm{x},\mathrm{y},\lambda)$ with respect to $x$, $y$, and $\lambda$

**(c)** (5 points) Set these partial derivatives equal to 0 and solve this set of three equations for the three unknowns $x$, $y$, and $\lambda$

**(d)** (5 points) Find the values of $x$, $y$, and $\lambda$ that maximize $f(x,y)$ subject to $g(\mathrm{x},\mathrm{y}) \geq 0$

**(e)** (5 points) Find the values of $x$, $y$, and $\lambda$ that maximize $f(x,y)$ subject to $g(\mathrm{x},\mathrm{y}) \leq 0$

**Problem 4.** (25 points) Consider the Boltzmann machine shown on the right with three units: two visible and one hidden. The units take on values $\{-1, +1\}$ and the weights between units are $w_{v_1 v_2} = \ln 3$ , $w_{v_1 h} = \ln 2$ , $w_{v_2 h} = \ln 1$




**(a)** (10 points) Write down, for each state (i.e., for all combinations of settings of the units), the expression for the energy, the un-normalized "probability", and the normalized probability at temperature $T = 1$.

**(b)** (5 points) Compute the probability that the visible units are in the state $(v_1, v_2) = (1,1)$ when the network is generating data freely (i.e., when the visible units are not clamped).

**(c)** (10 points) If the network is being trained on a single data point where the visible units are in the state $(v_1, v_2) = (1,1)$, what is the derivative of the log probability of the data with respect to $w_{v_2 h}$?

**Bonus Problem.** (10 points)

**(a)** (5 points) Demonstrate that the RBF, polynomial, and tanh kernels satisfy $k(Q\boldsymbol{x}, Q\boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}')$ for any matrix $Q$ that is unitary, i.e., $Q^{-1} = Q^T$

**(b)** (5 points) Prove whether or not this property holds for the following kernel? $k(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T A \boldsymbol{x}'$ where $A$ is a symmetric and positive semidefinite matrix.