CPSC 633 Exam #2 (4/25/2013, Friday)[1]

Last name: _____, First name: _____

Time: **10:20am–11:10am (50 minutes)**, Total Points: **100**

Subject	Score
Reinforcement Learning	/10
Genetic Algorithms	/10
Dimensionality Reduction	/20
Local Models	/20
Bayesian Learning	/20
Kernel Machines	/20
Total	/100

- Be as **succinct** as possible. Usually, one sentence is enough to answer one specific question. If your statement includes both correct and incorrect answers, you will get 0 points.
- Read the questions carefully to see what kind of answer is expected (*explain blah* **in terms of ...** *blah*).
- If you feel that the question is not specific enough, please do ask. Be careful not to mention any hint, or suggest alternative answers.
- This is a closed-book, open-note (the one-sheet hand-written note you brought) exam.

[1] Instructor: Yoonsuck Choe.

1 Reinforcement Learning

Question 1 (10 pts): Explain the following in terms of observed reward values and estimated reward values.

(1) What does $TD(\lambda)$ with $\lambda = 0$ correspond to? _____

(2) What does $TD(\lambda)$ with $\lambda = 1$ correspond to? _____

(3) Explain conceptually what $TD(\lambda)$ with $0 < \lambda < 1$ correspond to.

2 Genetic Algorithms

Question 2 (10 pts): Selection based on probatility that is proportional to the fitness can lead to crowding. (1) What are the two standard selection methods? (2) Explain how they can help reduce crowding.

(1) Answer: _____ and _____

(2) Explain:

3 Dimensionality Reduction

Question 3 (10 pts): (1) In PCA, can the eigenvalues tell you how much the data can be compressed? (2) Explain why or why not.

(1) Answer: _____

(2) Explain:

Question 4 (10 pts): (1) In Isomap, does the choice of the parameter ϵ affect the calculated geodesic distance between two arbitrary points in the data set? (2) Explain why or why not. (3) Once the geodesic distances are found, what algorithm is used for the final mapping?

(1) Answer: _____

(2) Explain:

(3) Answer: _____

4 Local Models

Question 5 (10 pts): (1) If the amount of information in the data is the same as the capacity of an information channel, can unsupervised learning be used successfully? (2) Explain in terms of redundancy in the data.

(1) Answer: _____

(2) Explain:

Question 6 (10 pts):

In RBF learning, in what situation do you need to use the pseudoinverse of the RBF layer activation matrix, not the inverse of it? Explain in terms of the relationship between the number of RBF units m and the number of input vectors n.

Answer: _____

5 Bayesian Learning

Question 7 (10 pts): (1) Does h_{MAP} always give the same result as a Bayes optimal classifier? (2) Explain why or why not.

(1) Answer: _____

(2) Explain:

Question 8 (10 pts): Explain how Occam's razor is related to Bayesian learning.

(1) What is the learning criterion that embodies Occam's razor?

Answer: _____

(2) How do the two main components of the learning criterion correspond to certain probabilities?

Explain:

6 Kernel Machines

Question 9 (10 pts):

(1) In SVM why is it important to maximize the margin ρ? Explain in terms of the VC dimension.
Explain:

(2) How does the margin ρ relate to the weight vector $\vec{w}$? Express the relation using a formula.

Answer: ρ = _____

Question 10 (10 pts): SVM usually projects the input vector into higher dimensional space. (1) What is the disadvantage in doing this (explain in terms of VC dimension)? (2) What is the advantage of doing this (explain in terms of a certain theorem). (3) How does SVM overcome the computational cost for projecting to higher dimensions with a nonlinear function? Explain in terms of a certain theorem.
(1) VC dimension (Increases / Decreases)

(2) Name the theorem (_____) and explain the main idea.
Explain:

(3) Name the thoerem (_____) and explain the main idea.
Explain: