YOUR NAME:
YOUR ID NUMBER:

SAMPLE MIDTERM FOR CS 2535, Spring 2004

Time: Friday Feb 27 2004, 11.10-12.10am

Answer ALL questions in Part A. Answer exactly 3 questions in part B.
Each question in part A is worth 3 points. Each question in part B is worth 5 points.
Each point is worth 1% of the final grade.

PART A (which will contain 5 questions)

Answer ALL questions in this part

1. Why does "early stopping" improve generalization?
2. When PCA is viewed from a Minimum Description Length perspective, which coding costs is it ignoring?

PART B (which will contain 5 or 6 questions)

Answer EXACTLY 3 questions in this part

1. a) Derive in detail the amount of information that must be communicated in order to send a value x to within a tolerance of t when the sender and receiver assume that x comes from a Gaussian distribution with standard deviation σ and mean μ.

   b) The minimum description length framework can be used to show that noisy weights in a supervised neural network are cheaper to communicate then precise ones. Explain how the receiver recovers the random bits of another message that were used to pick a value for a noisy weight from its posterior distribution?

2. Write down the expression for the energy gap $E(s_j = 1) - E(s_j = 0)$ where E is the energy of a configuration of a sigmoid belief net and j is a hidden unit that has two parents, h and i, and one descendant, k.