Machine Learning 2014-15
Final Exam

17 December 2014

Name: .....

Questions

Question 1: 2.5 points In linear regression, the in-sample error is given by

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\left(\mathbf{w}^{\text{T}}\text{X}^{\text{T}}\text{X}\mathbf{w} - 2\mathbf{w}^{\text{T}}\text{X}^{\text{T}}\mathbf{y} + \mathbf{y}^{\text{T}}\mathbf{y}\right),$$

where X is the $N \times d$ input matrix, $\mathbf{y}$ is the $N \times 1$ output vector, and $\mathbf{w}$ is the $d \times 1$ weight vector. Given that $\text{X}^{\text{T}}\text{X}\mathbf{w} = \text{X}^{\text{T}}\mathbf{y}$, show that the in-sample error can be written

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\mathbf{y}^{\text{T}}\left(I - \text{X}(\text{X}^{\text{T}}\text{X})^{-1}\text{X}^{\text{T}}\right)\mathbf{y}.$$

Question 2: 2.5 points You move into a new house; the phone is connected, and you're pretty sure that the phone number is 740511, but not as sure as you would like to be. As an experiment, you pick up the phone and dial 740511; you obtain a 'busy' signal. Are you now more sure of your phone number? If so, by how much? (Be as explicit as possible in your reasoning.)

Question 3: 1 points What are the components of supervised learning? Which components are under the control of someone that wants to apply supervised learning?

Question 4: 1 points What is data snooping, and what effect does it have on learning?

Question 5: 1 points Why is logistic regression harder than linear regression?

Question 6: 1 points How is the Vapnik-Chervonenkis dimension defined, and why is it useful?

Question 7: [1 points] What is the bias-variance tradeoff? Describe one way to decrease the bias and one way to decrease the variance.

Question 8: [1 points] When performing cross-validation, K points are set aside for testing. Which factors do you have to consider when choosing the value of K?

Question 9: [1 points] What does the backpropagation algorithm for neural networks estimate?

Question 10: [1 points] Describe a technique that diminishes the effect of overfitting in the specific case of learning decision trees.

Question 11: 1 points Describe what is meant by the "kernel trick". What property does an algorithm need in order to apply the kernel trick?

Question 12: 1 points An infinite-horizon MDP does not have goal states, i.e. execution never terminates and action selection keeps going indefinitely. If you are applying reinforcement learning to an infinite-horizon MDP, which reward measure would you choose, and why?

Question 13: 1 points When MDPs are too large, standard value-based methods are ineffective. Describe a reinforcement learning technique that is effective for large MDPs.

Question 14: 4 points Consider the hidden Markov model:

$$p(v_1, \ldots, v_T, h_1, \ldots, h_T) = p(h_1)p(v_1|h_1) \prod_{t=2}^{T} p(v_t|h_t)p(h_t|h_{t-1})$$

in which $\mathrm{dom}(h_t) = 1, \ldots, H$ and $\mathrm{dom}(v_t) = 1, \ldots, V$ for all $t = 1, \ldots, T$.

1. Draw a belief network representation of the above distribution.
2. Draw a factor graph representation of the above distribution.
3. Use the factor graph to derive a Sum-Product algorithm to compute the marginals $p(h_t|v_1, \ldots, v_T)$. Explain the sequence order of messages passed on your factor graph.
4. Explain how to compute $p(h_t, h_{t+1}|v_1, \ldots, v_T)$.

The total score is: _____ /20.