Machine Learning 2015-16
Final Exam

16 December 2015

Name: .....

Questions

Question 1: 2.5 points In linear regression, the *hat matrix* $\mathrm{H} = \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top$ is defined such that $\hat{\mathbf{y}} = \mathrm{H}\mathbf{y}$, where $\mathbf{y}$ are the true outputs and $\hat{\mathbf{y}}$ are the estimated outputs. Show that a) $\mathrm{H}^\top = \mathrm{H}$; b) $\mathrm{H}^2 = \mathrm{H}$; c) $(\mathrm{I}-\mathrm{H})^2 = (\mathrm{I}-\mathrm{H})$. Hints: $(\mathrm{AB})^\top = \mathrm{B}^\top\mathrm{A}^\top$ and $(\mathrm{A}^{-1})^\top = (\mathrm{A}^\top)^{-1}$.

Question 2: 2.5 points Late one night, a woman was mugged at the UPF Poblenou Campus. The perpetrator was masked and did not speak, but the woman claims it was another woman who robbed her. In tests under similar circumstances (same light level, same clothing, same distance), the woman correctly predicted the gender (male or female) of different people 80% of the time. Of the people working and studying at the Poblenou Campus, 70% are male and 30% are female. What is the probability that the perpetrator was a woman?

Question 3: 1 points The Hoeffding inequality applied to learning is

$$\mathbb{P}[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2e^{-2\epsilon^2 N}.$$

What is the meaning of this inequality, and why is it important for machine learning?

Question 4: 1 points Given a hypothesis set $\mathcal{H}$, how is a break point for $\mathcal{H}$ defined?

Question 5: 1 points What is the main difference between the union bound and the VC bound regarding the out-of-sample error?

Question 6: 1 points Which three factors do overfitting mainly depend on?

Question 7: 1 points Describe a technique that combats overfitting, including the steps necessary for the technique to function properly.

Question 8: 1 points Describe the steps necessary to classify a set of data points with binary output using support vector machines.

Question 9: 1 points Which are the different types of value-based methods for reinforcement learning, and how do they differ?

Question 10: 1 points Describe the backpropagation algorithm for feedforward neural networks. Which type of algorithm is backpropagation an example of?

Question 11: [1 points] Describe two different techniques for multiclass classification.

Question 12: [1 points] How does the error measure affect the final hypothesis of a machine learning algorithm?

Question 13: [5 points] The QMR-DT database is a bipartite graphical model in which the upper layer of nodes represent diseases and the lower layer of nodes represent symptoms. The evidence is a set of observed symptoms (findings) and represent the vector of findings with the symbol f. The symbol d denotes the vector of diseases. All nodes are binary, thus the components f_i and d_i are binary random variables.

1. Write down the joint probability model $p(f, d)$.
2. Draw the structure of the QMR-DT graphical model as a Bayesian network (represent the evidence as shaded nodes).
3. Draw the corresponding factor graph model of the QMR-DT Bayesian network.
4. Given the following "noisy-OR" model for the conditional probabilities:

$$P(f_i = 0|d) = (1 - q_{i0}) \prod_{j \in \pi(i)} (1 - q_{ij})^{d_j},$$

where q_{i0} is the probability that the ith finding is present even though no disease is present; and q_{ij} is the probability that the ith finding is present when disease jth is present (q_{i0} and q_{ij} are given by experts).

   (a) Suggest a graphical transformation that represents each factor involving n variables in terms of factors that involve at most 3 variables.

   (b) What is the main bottleneck for inference in this model? [hint: rewrite the noisy-OR model on an exponential form using $\theta_{ij} \equiv -\log(1 - q_{ij})$].

The total score is: _____ /20.