

# Machine Learning 2013-14

## Final Exam

12 December 2013

Name: .....NIA: .....

### Questions

**Question 1:** 2 points The hat matrix  $H = X(X^T X)^{-1} X^T$  expresses the in-sample error of linear regression such that  $\hat{y} = Hy$ , where  $y$  is the true output and  $\hat{y}$  is the estimated output. Show that  $H^T = H$  and that  $H^2 = H$ . Hint:  $(A^T)^T = A$ ,  $(AB)^T = B^T A^T$  and  $(A^{-1})^T = (A^T)^{-1}$ .

**Question 2:** 2 points Consider a sample of two marbles drawn independently from a bin that holds red and green marbles. The probability of a red marble is  $\mu = 0.3$ .

- a) If we draw one such sample of two marbles, what is the probability of drawing no red marbles?
- b) If we draw two samples, what is the probability of at least one sample containing no red marbles?

**Question 3:** 1 points The Hoeffding inequality applied to learning can be written as

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

What is the meaning of this inequality?

**Question 4:** 1 points How is learning affected when the output is non-deterministic? In other words, instead of estimating a target function  $y = f(x)$ , the aim is to estimate a conditional probability  $P(y | x)$ .

**Question 5:** 1 points If you had to classify a set of data points with binary output in  $d$  dimensions that you knew were linearly separable, which algorithm would you choose? What is the VC-dimension of the algorithm?

**Question 6:** 1 points If you had to classify a set of data points with binary output in  $d$  dimensions that you knew were not linearly separable, which algorithm would you choose? Motivate your choice.

**Question 7:** 1 points Given a hypothesis set  $\mathcal{H}$ , how is a break point for  $\mathcal{H}$  defined?

**Question 8:** 1 points What is the main difference between the union bound and the VC bound regarding the out-of-sample error?

**Question 9:** 1 points Which three factors do overfitting mainly depend on?

**Question 10:** 1 points Describe a technique that combats overfitting, including the steps necessary for the technique to function properly.

**Question 11:** 1 points What quantity do algorithms for learning decision trees try to optimize? What is the meaning of this quantity?

**Question 12:** 1 points Which clustering algorithm would you choose if the data points were arranged in (possibly overlapping) circular bands in two dimensions? Motivate your choice.

**Question 13:** 1 points Describe the steps necessary to classify a set of data points with binary output using support vector machines.

**Question 14:** 1 points Which are the different types of value-based methods for reinforcement learning, and how do they differ?

The total score is: \_\_\_\_\_ /16.