

# Machine Learning 2014-2015—Final Exam

27 March 2015, 15:15 – 18:15

The exam is **open book**: you can use Flach's *Machine Learning* (as a print of the PDF or as a proper book) as well as the lecture slides and any notes you've taken. You can use a calculator as long as it doesn't provide (internet) connectivity.

The exam grade determines 50% of your final grade for the course, the other half is determined by your project grade. The exam grade must at least be 5.5 to pass the course.

**Good luck!**

<i>Question</i>	<i>Points</i>
1. Short answers	15
2. Decision Trees	20
3. Bayesian Learning	15
4. Neural Networks	20
5. Instance-based Learning	15
6. Classification and Performance Measures	15
<b>Total</b>	<b>100</b>

## Questions

1. **Short answers** no justification required (*3 Points for each question*)

- (a) (**True** or **False**): As long as a decision tree only uses binary splits, it cannot overfit.
- (b) (**True** or **False**): Clustering techniques like k-means require some labeled data to initialise the cluster centres.
- (c) (**True** or **False**): When training neural networks, the network weights are updated more often in incremental or stochastic gradient descent than in batch gradient descent.
- (d) (**True** or **False**): Support Vector Machines can only model linear decision boundaries.
- (e) (**True** or **False**): Cross Validation is a method that can help select proper settings for machine learning algorithms.

## 2. Decision Trees

You're taking a date to the movies. To be sure that you both enjoy the night out, you choose to select a movie using a decision tree that you develop on data about movies you both watched. Table 1 contains the training data.

Genre	Oscar nominee	Lead	Liked by both?
Romantic Comedy	Yes	Male	No
Horror	Yes	Male	No
Romantic Comedy	No	Male	No
Arthouse	No	Male	No
Arthouse	No	Male	Yes
Romantic Comedy	No	Female	Yes
Horror	No	Male	Yes
Horror	Yes	Female	Yes

Table 1: Movies

- (a) (5 points) What are the entropies  $H(Liked|Genre = Arthouse)$  and  $H(Liked|Lead = Female)$ ?<sup>1</sup>
- (b) (5 points) The ID3 algorithm would select *Lead* as the root node of the tree. What will the algorithm split on next when *Lead = Male*? Show your calculations.
- (c) (5 points) What will the algorithm decide for the *Lead = Female* branch?
- (d) (5 points) Suppose that the decision tree is not elaborated further, i.e., there is no further split after the one you identified in question 2b. Draw the tree and calculate its training set error.

## 3. Bayesian Learning

- (a) (5 points) From the data in table 1, what is the prior probability of both you and your date liking a movie? Show your calculations.
- (b) (5 points) What would a naive Bayes classifier trained on that data predict for a romantic comedy with a male lead? Show your calculations.
- (c) (5 points) Would there be a difference between an MLE and an MAP prediction for a Bayes classifier trained on the data in Table 1? Why (not)?

---

<sup>1</sup>Note: if you need a calculator and yours can't do  $\log_2$ , use the following equation:  $\log_2(x) = 1.44 \cdot \ln(x)$ . If you didn't bring a calculator, you may give the answer as an expression.

#### 4. Neural Networks

In this question, we consider a neural network with two input nodes  $x_1$  and  $x_2$ , two hidden nodes  $h_1$  and  $h_2$  and one output node  $o$ . The nodes have sigmoid activation functions. The weights for the connections are as follows:

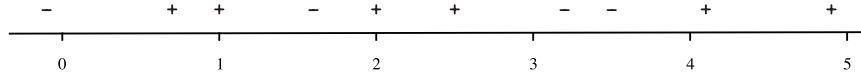
Connection	Weight
$x_1 \rightarrow h_1$	0.3
$x_1 \rightarrow h_2$	0.4
$x_2 \rightarrow h_1$	0.1
$x_2 \rightarrow h_2$	0.0
$h_1 \rightarrow o$	0.1
$h_2 \rightarrow o$	0.3

Table 2: Neural Network Weights

- (a) (5 points) Draw the network with the edges labelled with the weights.
- (b) (5 points) What is the output of the network if the input is (0.5,0.5)? Show your calculations.
- (c) (5 points) Let's assume that for these inputs, the target value  $t = 0.5$  and the learning rate  $\eta = 0.1$ . Compute the new weight for the connection  $h_1 \rightarrow o$  using backpropagation. Show your calculations.
- (d) (5 points) For a neural network, which one of the following structural assumptions most affects the trade-off between underfitting (i.e. low performance) and overfitting? Explain your reasoning.
  - i. The learning rate
  - ii. The initial choice of weights
  - iii. The number of hidden nodes
  - iv. The use of a constant-term unit input

## 5. Instance-based Learning

Consider the following dataset with one real-valued input  $x$  and one binary output  $y$ . We are going to use k-nearest neighbour with unweighted Euclidean distance to predict  $y$  for a given  $x$ .



The precise  $x$  values are as follows:  $\{-0.1; 0.7; 1.0; 1.6; 2.0; 2.5; 3.2; 3.5; 4.1; 4.9\}$

- (2.5 points) What is the predicted class of 1-nearest neighbour for a new data-point  $x = 0.0$ ?
- (2.5 points) What is the predicted class of 3-nearest neighbour for a new data-point  $x = 0.0$ ?
- (5 points) What is the leave-one-out cross-validation <sup>2</sup> error of 1-nearest neighbour on this dataset? Give your answer as the number of misclassifications.
- (5 points) Now, we will use a simple uniform kernel function that weights all data points that are at a distance of max. 1 equally, i.e.:

$$k(d) = \begin{cases} 1, & \text{if } d \leq 1. \\ 0, & \text{otherwise.} \end{cases}$$

What is the predicted class of nearest neighbour with this kernel for a new data-point  $x = 2.1$ ?

## 6. Classification and Performance Measures

- (5 points) For a binary classification problem, a classifier is evaluated on a dataset of 1000 cases. There are 700 positive and 300 negative examples in the dataset. It correctly identifies 586 positive cases and 234 negative cases. Draw the confusion matrix that follows from these numbers.
- (5 points) Give the precision, accuracy and recall for this classifier. Show your calculations.
- (5 points) Three classifiers are compared using cross-fold validation: the results on each of the four hold-out samples is reported in table 3. Which model should be selected? Show your calculations.

	Model 1	Model 2	Model 3
fold 1	0.87	0.89	0.90
fold 2	0.78	0.85	0.86
fold 3	0.79	0.84	0.76
fold 4	0.82	0.9	0.89

Table 3: Cross-fold validation results

---

<sup>2</sup>Remember: Leave-one-out cross-validation = cross-validation using each separate data-point as the hold-out sample once, i.e., first use  $\langle -0.1, + \rangle$  as the test set, then  $\langle 0.7, + \rangle$ , and so on.