Johannes Gutenberg-Universität Mainz
Institut für Informatik
Data Mining
Prof. Dr. Stefan Kramer

Machine Learning - Sheet 3

11.05.2017
Deadline: 18.05.2017 - 23:55

Task 1: Random Forests *(10 Points)*

In this task, you are supposed to compare the performance of decision trees with random forests on the car dataset (`http://archive.ics.uci.edu/ml/datasets/Car+Evaluation`).

(a) What is the relation between bagging and random forests? Briefly describe both methods, point out the differences.

(b) Make yourself familar with classification in WEKA (`http://www.cs.waikato.ac.nz/ml/weka/`) – either using the graphical user interface or the API.

(c) Analyze how the performance of WEKA's random forest is affected when using different numbers of trees (e.g., 1, 5, 10, 20, ...). To do so, use the car dataset and a very simple procedure: Randomly split the dataset into a training set and a test set (two thirds and one third). Take the first to train the random forest. Afterwards, use the test set to compute the percentage of correctly classified instances.

(d) Compare the performance of WEKA's random forests with your decision tree implementation.

Task 2: Probability Distributions On Finite Sets *(10 Points)*

Let Ω be a finite set. We call a function $\mathbb{P}[\{\bullet\}] : \Omega \to [0,1]$ a *probability mass function*, if it has the following property:

$$\sum_{\omega \in \Omega} \mathbb{P}[\{\omega\}] = 1.$$

We call elements ω of Ω *elementary events*, and we call subsets $A \subseteq \Omega$ *events*. We extend the notation to the (non-elementary) events $A \subseteq \Omega$ as follows:

$$\mathbb{P}[A] := \sum_{\omega \in A} \mathbb{P}[\{\omega\}].$$

Furthermore, for events $A, B \subseteq \Omega$, we define *conditional probability of A given B* as follows:

$$\mathbb{P}[A \,|\, B] := \begin{cases} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} & \text{if} \quad \mathbb{P}[B] > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Prove the following statements:

1. Let $n \in \mathbb{N}$ and $A_1, \ldots, A_n \subseteq \Omega$ be pairwise disjoint events. By *pairwise disjoint* we mean that for all $i \neq j$ the events A_i and A_j are disjoint, that is, $A_i \cap A_j = \emptyset$. We write $\biguplus_{i=1}^{n} A_i$ for $\bigcup_{i=1}^{n} A_i$ to emphasize that it is the union of disjoint sets.

   Prove that $\mathbb{P}[\biguplus_{i=1}^{n} A_i] = \sum_{i=1}^{n} \mathbb{P}[A_i]$.

   Hint: It's a single chain of equations, where each step follows immediately from the above definitions.

2. Let $A, B \subseteq \Omega$ be two (not necessarily disjoint) events. Show the *sum rule*:

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

   Hint: Subdivide $A \cup B$ into three pairwise disjoint sets. Draw a Venn diagram, if necessary.

3. Show that for all events A it holds: $\mathbb{P}[A \mid \Omega] = \mathbb{P}[A]$.

4. Let $n \in \mathbb{N}$ and for $i = 1, \ldots, n$ let $H_i \subseteq \Omega$ be such that $\biguplus_{i=1}^{n} H_i = \Omega$. Let A be some event. Show the *law of total probability*:

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \mid H_i] \, \mathbb{P}[H_i].$$

   Hint: Use the fact that $A = A \cap \Omega$, apply distributivity of $\cap$ over $\cup$.

5. Let A, B be two events with $\mathbb{P}[B] > 0$. Prove the *Bayes' theorem*:

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A] \, \mathbb{P}[A]}{\mathbb{P}[B]}.$$

6. Let $B, H_1, \ldots, H_n$ be such that $\biguplus_{i=1}^{n} H_i = \Omega$. Show that for each $j = 1 \ldots n$ it holds:

$$\mathbb{P}[H_j \mid B] = \frac{\mathbb{P}[B \mid H_j] \, \mathbb{P}[H_j]}{\sum_{i=1}^{n} \mathbb{P}[B \mid H_i] \, \mathbb{P}[H_i]}.$$

   Hint: Combine the basic version of Bayes' theorem and the law of total probability.

7. Let $A_1, \ldots, A_n$ be some events. Show the *chain rule of conditional probabilities*:

$$\mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \mathbb{P}\left[A_i \,\middle|\, \bigcap_{j=1}^{i-1} A_j\right].$$

   Hint: Split $\bigcap_{i=1}^{n} A_i$ into $A_n \cap \bigcap_{i=1}^{n-1} A_i$, apply induction over n; $\bigcap_{j=1}^{0} A_j = \Omega$.