Exercises to the lecture

Machine learning and Pervasive Computing

Assignment 1: Data sampling and test design

In this assignment, you are asked to decide for a class to detect/distinguish from signal-strength information, to collect data, to train a classifier on the data and to compute statistics on the performance of the classifier. The class to be detected can be chosen according to your preference. For this,

   a) collect RSSI information from a WiFi interface (e.g. utilising tcpdump or airodump-ng).

   b) Then, following one of the test designs discussed in the lecture, create separate training and testing sets.

   c) Finally, utilising a data mining toolkit (e.g. WEKA or Orange), train a classifier of your choice with the data and

   d) compute the classification performance utilising one of the metrics introduced in the lecture (e.g. IS, Brier, AUC, F_1).

Assignment notes

- Assignments are designed to be completed in groups. Please coordinate and distribute the work within your group and discuss actions to take and problems encountered.

- In order to access the RSSI information, each group is provided with a mobile phone with the possibility to capture such information. However, RSSI information captured with other devices might be sufficient as well.

- In some cases it might be necessary to install a different firmware on the Nexus One phones in order to be able to access the wireless channel in monitor mode[1]. The firmware is available via http://bcmon.blogspot.de/.
  (there is also an apk: https://bcmon.googlecode.com/files/bcmon.apk and an app from a prior advanced practical course which would come with the firmware: `git checkout https://github.com/crauterb/pcan_Analyzer.git`)

- After installation, the phone should have an additional wireless interface in monitor mode (named eth0; please confirm with `iwconfig` on the phone (for instance via `adb shell`))

- Using `tcpdump` on the phone, RSSI of packets at the wireless interface can be monitored (e.g. `tcpdump > <filename>` or `tcpdump -w <filename>`; Note that write access might be possible only to `/sdcard`)

- For easy classification and performance evaluation, standard data-mining toolkits should be used (for instance, Weka or Orange)[2].

- Note that it might be necessary to utilise packets from one WiFi AP only (since a second AP might have different base-RSSI level).

---

[1] monitor mode is necessary in order to obtain RSSI information from all packets on the channel
[2] Orange: http://orange.biolab.si/; Weka: http://www.cs.waikato.ac.nz/ml/weka/