# 732A95 INTRODUCTION MACHINE LEARNING

# LAB 5: KERNEL METHODS AND SUPPORT VECTOR MACHINES

JOSE M. PEÑA
IDA, LINKÖPING UNIVERSITY, SWEDEN

## INSTRUCTIONS

Each student must submit a report with his/her solutions to the lab. Submission is done via LISAM and before the deadline. The submission file should be named Name_LastName.pdf. The report must be concise but complete. It should include (i) the code implemented or the calls made to existing functions, (ii) the results of such code or calls, and (iii) explanations for (i) and (ii).

PhD students pass the lab if their individual report is of sufficient quality. TDDE01 and 732A95 students pass the lab as follows. The students must discuss their lab solutions in a group. Each group must compile a collaborative report that will be used for presentation at the seminar. The report should clearly state the names of the students that participated in its compilation and a short description of how each student contributed to the report. This report should be submitted via LISAM and before the deadline. The file should be named Group_X.pdf where X is the group number. The collaborative reports are corrected and graded. The individual reports are also checked, but feedback on them will not be given. The students pass the lab if their group report passes the seminar and their individual reports have reasonable quality, otherwise the students must complete their individual reports by correcting the mistakes in them.

## RESOURCES

**Please use only basic R functions in your solution, e.g. sum() and alike**. The only exception is that you are allowed to use `distHaversine` from the `geosphere` package.

## ASSIGNMENT

Implement a kernel method to predict the hourly temperatures for a date and place in Sweden. To do so, you are provided with the files `stations.csv` and `temps50k.csv`. These files contain information about weather stations and temperature measurements for the stations at different days and times. The data have been kindly provided by the Swedish Meteorological and Hydrological Institute (SMHI) and processed by Zlatan Dragisic.

You are asked to provide a temperature forecast for a date and place in Sweden. The forecast should consist of the predicted temperatures from 4 am to 24 pm in an interval of 2 hours. Use a kernel that is the sum of three Gaussian kernels:

- The first to account for the distance from a station to the point of interest.
- The second to account for the distance between the day a temperature measurement was made and the day of interest.
- The third to account for the distance between the hour of the day a temperature measurement was made and the hour of interest.

Choose an appropriate smoothing coefficient or width for each of the three kernels above. Answer to the following questions:

- Show that your choice for the kernels' width is sensible, i.e. that it gives more weight to closer points. Discuss why your of definition of closeness is reasonable.
- It is quite likely that the predicted temperatures are too low. Do you think that the reason may be that the three Gaussian kernels are independent one of another ?

Note that the file `temps50k.csv` may contain temperature measurements that are posterior to the day and hour of your forecast. You must filter such measurements out, i.e. they cannot be used to compute the forecast. Feel free to use the template below to solve the assignment.

```
set.seed(1234567890)
library(geosphere)

stations <- read.csv("stations.csv")
temps <- read.csv("temps50k.csv")
st <- merge(stations,temps,by="station_number")

h_distance <- # These three values are up to the students
h_date <-
h_time <-
a <- 58.4274 # The point to predict (up to the students)
b <- 14.826
date <- "2013-11-04" # The date to predict (up to the students)
times <- c("04:00:00", "06:00:00", ..., "24:00:00")

temp <- vector(length=length(times))

# Students' code here

plot(temp, type="o")
```