# Machine Learning
# Exercise 4

Marc Toussaint
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

May 12, 2015

## 1 Labelling a time series

Assumes we are in a clinical setting. Several sensors measure different things of a patient, like heart beat rate, blood pressure, EEG signals. These measurements form a vector $x_t \in \mathbb{R}^n$ for every time step $t = 1, .., T$.

A doctor wants to annotate such time series with a binary label that indicates whether the patient was asleep. Your job is to develop a ML algorithm to automatically annotate such data. For this assume that the doctor has already annotated several time series of different patients: you are given the training data set

$$D = \{(x^i_{1:T}, y^i_{1:T})\}_{i=1}^n \ , \quad x^i_t \in \mathbb{R}^n,\ y^i_t \in \{0,1\}$$

where the superscript $i$ denotes the data instance, and the subscript $t$ the time step (e.g., each step could correspond to 10 seconds).

Develop a ML algorithms to do this job. First specify the model formally in all details (representation & objective). Then detail the algorithm to compute optimal parameters in pseudo code, being as precise as possible in all respects. Finally, don't forget to sketch an algorithm to actually do the annotation given an input $x^i_{1:T}$ and optimized parameters. (No need to actually implement.)

## 2 CRFs and logistic regression

Slide 03:26 summarizes the core equations for CRFs.

a) Confirm the given equations for $\nabla_\beta Z(x,\beta)$ and $\nabla^2{}_\beta Z(x,\beta)$ (i.e., derive them from the definition of $Z(x,\beta)$).

b) Derive from these CRF equations the special case of logistic regression. That is, show that the gradient and Hessian given on slide 03:11 can be derived from the general expressions for $\nabla_\beta Z(x,\beta)$ and $\nabla^2{}_\beta Z(x,\beta)$. (The same is true for the multi-class case on slide 03:17.)

## 3 Kernel Ridge regression

In exercise 2 we implemented Ridge regression. Modify the code to implement Kernel ridge regression (see slide 03:31). Note that this computes optimal "parameters" $\alpha = (K + \lambda I)^{-1} y$ such that $f(x) = \kappa(x)^\top \alpha$.

a) Using a linear kernel (see slide 03:34), does this reproduce the linear regression we looked at in exercise 2? Test this on the data. If not, how can you make it equivalent?

b) Is using the squared exponential kernel $k(x, x') = \exp(-\gamma \,|\, x - x' \,|\,^2)$ exactly equivalent to using the radial basis function features we introduced?