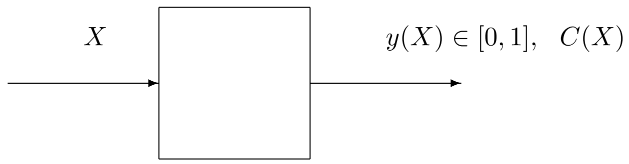
Master 2 Studentship at Lab. I3S

Title: Estimation of Risk of Decompression Sickness
Period (minimum): March 15 – July 15, 2012
Funding: $\approx$ 600€/month (by BF-Systèmes)
Supervision:

Luc Pronzato (pronzato@i3s.unice.fr)
Joao Rendas (rendas@i3s.unice.fr)
Julien Hugon (julien.hugon@bf-systemes.fr)

Description: The goal of this studentship is to establish a model of the risk of accidents of decompression sickness. More precisely, given a set of *input* parameters $X \in \mathbb{R}^d$ we want to compute the associated risk of accident $y(X) \in [0, 1]$.



If we let the binary random variable $DCS(X) \in \{0,1\}$ represent the occurrence of an accident (DCS = decompression sickness) for a dive characterized by X, then

$$y(X) = \text{Prob}\{DCS(X) = 1\} \quad .$$

Besides predicting the risk, we need to assess the confidence on the estimated $y(X)$, ideally in a form of a confidence interval $C(X)$ or a standard deviation.

A more complex version of the problem would try to model the occurrence of s different symptoms of DCS jointly, with now $y_i(X) = \text{Prob}\{DCS_i(X) = 1\}$, $i = 1, \dots, s$.

The map $X \longrightarrow y(X)$ will be learned from existing datasets. Tens of thousands of exposures characterized by
(i) bottom depth/duration/mixed gas breathed,
(ii) decompression procedure,
will be available, with the indication of all observed occurrences of DCS.

Several difficulties can be foreseen.

1. The input space is not uniformly sampled. We believe that it expresses the following priors: *(i)* the safer decompression procedures are preferred; *(ii)* no samples at values of risk higher than a threshold γ are available;
2. The number of accidents recorded is small (rare event), which must be taken into account in the construction of a procedure aimed at ensuring good confidence on its predictions.

Two approaches will be considered for the construction of the models.

- Neural network, which are known to be estimators of posterior probabilities in classification problems (with overlapping classes) if trained with a convenient error function, see, e.g., Miller et al. (1993). The problem maybe that the standard error functions do not behave well under unbalanced datasets, where one class is much more present than the other. Modified error functions have been defined to compensate for this. It is not clear how the confidence of the estimates is computed (this is the problem of recursive learning in NNs). Also, the choice of network topology is critical, and formal tools for that need to be found.

- Prediction by Kriging: this consists in learning $y(X)$ by a regression in other variables using the Gaussian Process (GP) model. Additional difficulties compared to the usual framework come from the fact that the predictions should be constrained to lie in $[0, 1]$ and that observations (occurrences of accidents) are heteroscedastic random variables.

The application to real data provided by BF-Systèmes will be considered.

References

Miller, J., Goodman, R., Smyth, P., 1993. On loss functions which minimize to conditional expected values and posterior probabilities. IEEE Transactions on Information Theory 39 (4), 1404–1408.