Kernel Logistic Regression for Phoneme Recognition

Peter Karsmakers*,**, Kristiaan Pelckmans**, J.A.K. Suykens**, B. De Moor**
*Katholieke Hogeschool Kempen (Associatie KULeuven), Kleinhoefstraat 4, 2440 Geel, Belgium
**Dep. of Electrical Engineering, SCD/SISTA, K.U.Leuven, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
Email: peter.karsmakers@esat.kuleuven.be

This research studies the extension of a multiclass logistic regression technique for the task of phoneme recognition. Herefor, a kernel version is derived based on a penalized likelihood criterion. The choice of this approach over an empirical risk minimization approach as performed by the Support Vector Machines (SVMs), is that the former yields probabilistic outcomes instead of a binary decision. This is particularly important in this subtask of speech recognition as it permits a proper integration of the phoneme recognition module in the full sequence. Specifically, it allows for a proper connection to a Hidden Markov Model (HMM) which makes different words out of a sequence of phonemes.

Kernel Logistic Regression

The training set can be written as $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \{1, \ldots, L\}$ where there are $L \geq 2$ different classes observed, together with d different covariates for any sample. Multiclass logistic regression starts from a stochastic model for each class which can be written as

$$\begin{cases} P(Y=1 \mid X=x) = \frac{1}{1+\sum_{l=2}^{L}\exp(\beta_l^T x)} \\ P(Y=2 \mid X=x) = \frac{\exp(\beta_2^T x)}{1+\sum_{l=2}^{L}\exp(\beta_l^T x)} \\ \vdots \\ P(Y=L \mid X=x) = \frac{\exp(\beta_L^T x)}{1+\sum_{l=2}^{L}\exp(\beta_l^T x)} \end{cases} \tag{1}$$

Here, $\beta_2, \ldots, \beta_L \in \mathbb{R}^d$ denote the parameters of the different models, while the first term ensures the normalization of the L different models. The common method to infer the parameters $\{\beta_l\}_{l=2}^L$ is via the use of penalized maximum likelihood which can be written as

$$\mathscr{L}_\gamma(\beta) = \prod_{i=1}^{n} P(Y=Y_l|X=x_i) \prod_{l=2}^{L} (\beta_l^T \beta_l)^\gamma. \tag{2}$$

Most often, a Newton-Raphson based strategy is used to optimize the loglikelihood. It is well-known that this procedure can be rewritten in terms of an iteratively reweighted least squares (IRLS) algorithm which consists of two steps: compute $\Delta\beta$ by solving a linear system and recompute weights.

Here we study the nonlinear extension to kernel machines where the inputs x are mapped to a high dimensional space via $\varphi(\cdot)$. Now, both steps can be easily reformulated in terms of a weighted LS-SVM.

A Robust Large Scale Algorithm

Succesful application of logistic regression depends crucially on a robust algorithm for maximizing the penalized likelihood function. We propose three modifications to the classical scheme:

- We extend the IRLS algorithm in order to incorporate the global regularization term. It turns out that this term is reflected in each step by a local regularization term, resulting in a robust algorithm when λ is chosen appropriately.
- While IRLS performs a Newton-Raphson optimization strategy, we resort to a version where one descends per parameterset β_l. Though the convergence becomes slightly worse, each step can be calculated much faster when using the dual representation.
- Since the size of the Hessian in the dual form is still proportional to the number of datapoints, we suggest an approach where each optimizationstep is based on a well-chosen subset of the available dataset.

Application to Phoneme Recognition

Experiments are carried out on the TIMIT dataset which consists of quasi-phonetically balanced American English training sentences, segmented on the phoneme level.

References

[1] Hastie T., Tibshirani R., Friedman J., "The Elements of Statistical Learning", *Elsevier*, 2001.

[2] A. Ganapathiraju, "Support vector machines for speech recognition", *PhD thesis*, Mississipi State University, 2002.

[3] J.A.K. Suykens, T. van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, "Least Squares Support Vector Machines", *World Scientific*, Singapore, 2002.

[4] Pelckmans K., "Primal-Dual Kernel Machines", *PhD thesis*, Faculty of Engineering, K.U.Leuven (Leuven, Belgium), 2005

Acknowledgements Research supported by BOF PDM/05/161 (Postdoc mandaat). (Research Council KUL): GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering, several PhD/postdoc & fellow grants; (Flemish Government): (FWO): PhD/postdoc grants, projects, G.0407.02, G.0197.02, G.0141.03, G.0491.03, G.0120.03, G.0452.04, G.0499.04, G.0211.05, G.0226.06, G.0321.06, G.0553.06, research communities (ICCoS, ANMMM, MLDM); (IWT): PhD Grants,GBOU (McKnow), Eureka-Flite2 - Belgian Federal Science Policy Office: IUAP P5/22,PODO-II,- EU: FP5-Quprodis; ERNSI; - Contract Research/agreements: ISMC/IPCOS, Data4s, TML, Elia, LMS, Mastercard JS is a professor and BDM is a full professor at K.U.Leuven Belgium, respectively.