Progress Week 7

Mentors:
Alexandre Gramfort, Marine Le Morvan
Intern:
Manon Rivoire
manon.rivoire@inria.fr

INRIA - Parietal Team

June 25, 2020

Overview

Lasso solver

- Implementation ok for sparse and dense,
- Issue concerning the binning : difference of shape between the one of y and the one of $X\beta$ corresponding to the value predicted by the Lasso model,
- Issue concerning the reading of the dataset *housing prices* (issue about the format of the dataset after pre-processing).
- Values of cross-validation score obtain with our Lasso solver are not the same as the one obtained with sklearn :
  originaldenselassocrossvalscore : -0.2548772178307937
  binninglassocrossvalscore : -1.0
  sklearndenselassocrossvalscore : 0.9982389866338556
  sklearnbinninglassocrossvalscore : 0.9820906240624006

Safe Pattern Pruning

The goal of the safe pattern pruning method is to easily discard groups of patterns which are guaranteed not to take part to the final optimal predictive model. For this purpose, the idea is to build a tree over the whole set of patterns and to discard entire subtrees according to a safe pattern pruning criterion $SPPC(t)$ which provides a condition at the level of the root node t asserting that we can discard the whole subtree depending on the root node t being sure that the patterns in this subtree are irrelevant for the final predictive model.

- Primal - Dual problems, minimization with regards two variables ? Expression of the L1-norm with regards (w, b) ? How to find the expression of the general dual problem given at page 3 equation (5) ?
- How to determine the inequality constraints related to the dual problem ?
- How to prove the two implications derived from the node inclusions just below the dual problem ?
- Lemma 7 ?

Safe Pattern Pruning (cont.)

- SPP Algorithm
- Main difference between screening and SPP ? Screening only allows to discard one feature by one feature whereas safe pattern pruning is able to discard entire groups of features ?
- Main difference between the SPP method and the boosting-type method only relies in the fact that the SPP method carry out a single search over the database whereas the boosting-type method carry out multiple searches over the database ? Is it the reason why safe feature screening is intractable for pattern mining task ?
- I don't understand the process consisting to add one violating constraint at a time in boosting-type method ?
- I don't understand very well how SPP is applied to graph mining and item set-mining
- Why is two-stage approach suboptimal ?

Safe Pattern Pruning (cont.)

- I do not understand how does the direct approach work concretely ? Why do we run an optimization step at each time that we add a new feature to the already selected features ?
- Is sparse modeling always a convex problem ?
- The selected set of features after pruning is larger than the optimal set of features ? Why ?
- How to prove in lemma one that the optimal solution only depends on the active feature ? (end of page 3)
- Theorem 2 (page 4) : What do the amounts u_t and v_t concretely represent ?
- I do not understand how to obtain the equations 9 and 10 in the proof of corollary 3
- Lemma 4 is derived from the KKT conditions but how to obtain $sign(u_t^*)$ instead of $sign(w_j)$ and $[-1, 1]$ instead of $[-\lambda, \lambda]$?

Safe Pattern Pruning (cont.)

- How to prove the lemma 5 (page 4) ? How to use the support function for this purpose ?
- How to obtain equation 7 from lemma 5 ?
- Proof of lemma 6 (page 5) : How to obtain equation 13 and the 4 KKT conditions ?
- Poof of lemma 7
- How to obtain the expression of λ_{max} ? (page 5) as well as the expression of its upper bound ?

SPP Algorithm

- Do we have to transform the database that we have into a graph, specifically a tree based on the features ?
- Do we have to implement SPP rules in separated functions ? The upper bound and SPPC ? The building of the tree based on the set of features composing the database ? The scan of the tree from the root to the leaves ? (depth first search ?)
- To solve the optimization problem restricted to the superset $\hat{\mathcal{A}}$ do we have to use the sparse lasso solver ?
- Do we have to return all the optimal solutions at each iteration ?
- How to integrate binning in the SPP method and what is the aim ?

Perspectives

- Implementation of SPP algorithm
- Tests of the solver with and without binning on the different datasets
- Summary of the obtained results
- Finish proofs of the SPP paper
- Block of lasso solvers in dense and sparse, block of SPP algorithm
- How to gather both lasso solvers and SPP ?
- Do we have to replace screening by SPP ?
- How to reach the final goal with these different blocks ?

Bibliography

- Safe Pattern Pruning Paper