

Machine Learning & Optimization for Cancer Research

This post contains several thesis proposals, all connected with an ambitious research project we have recently been assigned through a 5-year grant obtained from AIRC, the Italian Association for Cancer Research.



Possible thesis topics include:

- How to deal with missing data in ML approaches applied to patient data? Recall that data might be missing as a consequence of a choice (not prescribing an analysis), so that missing data might be semantically significant
- How to deal with inhomogeneous data in the training set? Some patients have been analyzed several times, some other just once. How to deal with this difference in the available features?
- Promoting interpretability in ML for patient data: in a context like this one, it is preferable to leave a few points of accuracy out, if this lets us develop a transparent and interpretable decision model. Decision trees and Risk Scores are two examples of interpretable

ML tools. Building accurate models reduces to a hard optimization problem

These, and other future, theses will profit of the availability of data made available to the project by the Careggi Hospital and the other partners of the Research Project