Project proposal: medical prediction about kidney disease

|  |  |
| --- | --- |
| |  | | --- | | *•* | |

**Stage 2** (10 points): Formulate your problem. In your **stage2** folder, put your report with the motivations and problem definitions. Note that this report can be updated in later stages.

Predicting whether a patient carrying a certain disease is a challenging data mining / machine

learning task since it has been traditionally performed by well-trained healthcare professionals.

For that reason, it requires expert knowledge in the particular field the data is sourced.

Furthermore, correlations within health record datasets are often very subtle that may be

undetectable, especially in smaller datasets. Therefore, training a “machine” for the task of

differentiating between patients and non-patients can become a very complex data mining /

machine-learning dilemma.

We had access to kidney diseases dataset at the UC-Irvine Machine Learning Repository. This

opened to us the possibility of exploring the challenges of testing different algorithms (one of

which we devised) for predicting the presence of diseases