Visual machine learning in spreadsheets
Advait Sarkar, Mateja Jamnik, Alan Blackwell, & Martin Spott

Spreadsheets are a natural fit for machine learning

We augment the spreadsheet, a versatile yet familiar data manipulation tool, to enable non-expert end-users to build and apply machine learning models.

The user selects rows of complete, correct data to “teach” the software, which implicitly trains a statistical model. The user can then apply this model to guess values for empty cells, or validate the contents of filled cells.

Visualisations improve end-user engagement

Through visualisations of the model along with meta-information such as class representation, the user can perform more nuanced activities, like judging the quality of the model, identifying good training examples, and querying why and how the model makes certain predictions.

Further reading

CLASS REPRESENTATION
These graphs show how the distribution of taught data differs from that of the overall distributions of data in the spreadsheet. Here, Iris-versicolor appears to be under-represented.