Presentation Notes

Algorithmic Assessment

Previously: We have seen that human assessment can keep up with COMPAS where just 7 features have been provided, whereas COMPAS uses 137 inputs.

Now: Can the accuracy of COMPAS (whose background functionality is unknown) be achieved with simple and interpretable classifiers?

Two methods:

- Logistic regression (with only small number of covariates)
- Nonlinear support vector machines

Dataset:

- Full database with more than 7000 defendants
- A model is trained on the 80% training set and its accuracy then determined by application to the 20% test set

Logistic Regression:

- Linear Classifier!
- Uses logistic functions to convert log-odds to probabilities

Feature Selection:

- Same 7 features as human assessment
- Only 2 features which are most predictive: age and total number of previous convicts

Support vector machine:

- Another classifier by Vladimir Vapnik and his colleagues
- Also separates data into two classes as LR
- Where the separating hyperplane is constructed such that it maximized distance between the 2 categories
- By transforming cartesian space where the covariates live with a kernel function one can build separators which are not just linear hyperplanes
- In this case radial basis kernel with hyperparameter θ is used which is an exponential with the negative distance of two points in argument

! Previous only linear separation, but now more sophisticated separations possible

> Now 3 methods of simple models at hand to compare to each other

Accuracy as criterion as before:

> Comparison is generated by 1000 times splitting, training, testing and averaging the accuracy Explain plot

Clear: they are all pretty much the same in terms of accuracy ! More features do not improve result ! Non-linear separation does not improve results ! Comparison with nonlinear method ensures that linear models do not limit performance (but performance is limited by data)

For the particular measure of fairness:

> Explain plot

- NL-SVM stands out that in general less FP on the cost of more FN
- But no method is racially fair here

Conclusion for the Algo Assess: They used 3 rather simple models to compare with COMPAS in terms of accuracy and fairness:

- simple model can perform as good as COMPAS (only 2 covariates instead of 137 and linear!)
- linear model does not restrict the performance (what we can see from the NL-SVM)

Comments apply for all commerical recidivism softwares. We have seen: COMPAS not more reliable than non-expert humans or very simple classifiers However, results demand for further TECHNICAL comments:

- Usage of particular measure for accuracy and fairness, but there are others...
- Usage of data on humans, they have a free mind and live in a biased world... Further analysis:
- What are the cases when the classifications disagree? Are there clear and non-clear cases? E.g. are only FP in cat 5 or also in cat 10?
- Most important: In the here explicated problem setting - the data seems not separable into to classes!

defendants! recidivism!