Topic 5: Tuning support vector machines

SVM specification and tuning

Categorical data can be handled by introducing binary *dummy* variables to indicate each possible value.

When fitting an SVM, the user must specify some control parameters, these include cost constant C for slack variables, the type of kernel function, and its parameters. Unlike the more probabilistic forms of classification, it is difficult to predict the out-of-sample classification error for SVMs, so cross-validation is used.

The following kernel functions available via the `R e1071` package:

linear: $\boldsymbol{u}^\top \boldsymbol{v}$

polybomial: $(\gamma \boldsymbol{u}^\top \boldsymbol{v} + c_0)^p$

radial basis: $\exp(-\gamma \|\boldsymbol{u} - \boldsymbol{v}\|^2)$

sigmoid: $\tanh(-\gamma \boldsymbol{u}^\top \boldsymbol{v} + c_0)$

for constants γ, p, and c_0.

Conclusion

We conclude this week with a brief discussion of the advantages and disadvantages of SVMs. SVM training can be formulated as a convex optimisation problem, with efficient algorithms for finding the global minimum, and the final result involves support vectors rather than the whole training set. This is both a computational benefit, but also one to robustness: outliers have less effect than for other methods.

On the other hand, they are much more difficult to interpret than modelbased classification techniques like the linear discriminant analysis. Furthermore, SVMs do not actually provide class probability estimates. These can be estimated by cross-validation, however.

To conclude the week, complete the demonstration and associated challenge task in the following slides.

Demonstration: Support vector machines

Please watch the following demonstration by Dr Krivitsky before proceeding to the task.

[Transcript](#)

This demonstration can be completed using the provided RStudio or your own RStudio.

To complete this task select the 'Cancor_Examples.demo.Rmd' in the 'Files' section of RStudio. Follow the demonstration contained within the RMD file.

If you choose to complete the example in your own RStudio, upload the following file:

📄 SVM_Examples.demo.Rmd

The output of the RMD file is also displayed below:

Challenge: Support vector machines

If you choose to complete this task in your own RStudio, upload the following file:

📄 SVM_Examples.challenge.Rmd

Click on the 'SVM_Examples.challenge.Rmd' in the 'Files' section to begin. Enter your response to the tasks in the 'Enter your code here' section.

This activity and the solution will be discussed at the Collaborate session this week. In the meantime, share and discuss your results in the 'Tutorials' discussion forum.

The solution will also be available here on Friday of this week by clicking on the 'Solution' tab in the top right corner.

The output of the RMD file is also displayed below: