Subsection 4

Boosting

Many views and aggregation

In bagging/RF (regression):

- many views are different samples
- aggregation is average

Alternative:

- many views are subsequent residuals
- aggregation is the sum

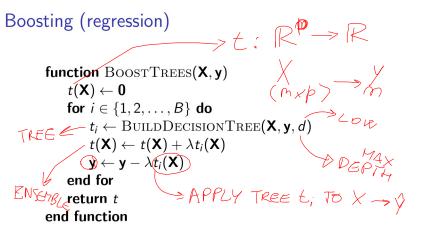
Boosting

When learning:

- 1. Current data is learning data
- 2. Repeat B times

When predicting:

- 1. Repeat B times
 - 1.1 get a prediction from *i*th learned tree
- 2. sum prediction
- Q: implementation differences w.r.t. RF?



- ► Each learned tree should be simple (maximum splits d)
- $ightharpoonup \lambda$ slows down learning

Trickier with classification.

Boosting parameters

- \blacktriangleright λ usually set to 0.01 or 0.001
- \triangleright λ and B interact: for small λ , B should be large
- ▶ large B can lead to overfitting (unlike bagging/RF, Q: why)

Find a good value for B with cross-validation

(Both boosting and bagging general techniques)

Bagging/RF/boosting in summary

	Tree	Bagging	RF	Boosting
interpretability	A			
numeric/categorical	A	A	A	A
accuracy	▼		A	A
test error estimate		A	A	
variable importance		A	A	A
confidence/tunability		A	A	
fast to learn	^ *			▼
(almost) non-parametric		A	A	
* • • • • • • • • • • • • • • • • • • •	1			

^{*:} Q: how faster? when? does it matter?

Lab: visualize forest errors and boundaries (3 h)

Consider just versicolor and virginica and their classification:

- 1. investigate variable importance with RF
 - 1.1 verify that it is true by removing important variables
- 2. investigate influence of B (ntree)
- 3. compare against decision tree
- 4. plot fuzzy decision boundaries

Packages: randomForest

Functions: geom_raster

>GGPLOT9

> meon Decresse Gir