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
   - 2.1 learn a tree on current data
   - 2.2 current data becomes residuals of learned tree $(\mathbf{y} - \hat{\mathbf{y}})$

When predicting:

1. Repeat B times
   - 1.1 get a prediction from ith learned tree
2. sum prediction

Q: implementation differences w.r.t. RF?

Boosting (regression)

function BOOSTTREES($\mathbf{X}, \mathbf{y}$)
  $t(\mathbf{X}) \leftarrow \mathbf{0}$
  for $i \in \{1, 2, \ldots, B\}$ **do**
    $t_i \leftarrow$ BUILDDECISIONTREE($\mathbf{X}, \mathbf{y}, d$)
    $t(\mathbf{X}) \leftarrow t(\mathbf{X}) + \lambda t_i(\mathbf{X})$
    $\mathbf{y} \leftarrow \mathbf{y} - \lambda t_i(\mathbf{X})$
  end for
  return t
end function

- Each learned tree should be simple (maximum splits d)
- λ slows down learning

Trickier with classification.

Boosting parameters

- λ usually set to 0.01 or 0.001
- 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	▲			
numeric/categorical	▲	▲	▲	▲
accuracy	▼		▲	▲
test error estimate		▲	▲	
variable importance		▲	▲	▲
confidence/tunability		▲	▲	
fast to learn	▲*			▼
(almost) non-parametric		▲	▲	

*: **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`