Today

- Ensemble methods
  - Bagging
  - Random forest
  - Boosting

Boosted regression tree

Algorithm

load y, x, x_{new}

set parameters: d, ntrees, λ

set $\hat{f}(x_{\mathrm{new}})$ = 0

set $r \leftarrow y$ (residuals equal to the data)

for b in 1 to ntrees

  train d split tree model on r and x

  predict residuals, $\hat{r}_b(x)$, from trained tree

  update residuals: $r \leftarrow r - \lambda \hat{r}_b(x)$

  predict y increment, $\hat{f}_b(x_{\mathrm{new}})$, from trained tree

  update prediction: $\hat{f}(x_{\mathrm{new}}) \leftarrow \hat{f}(x_{\mathrm{new}}) + \lambda \hat{f}_b(x_{\mathrm{new}})$

return $\hat{f}(x_{\mathrm{new}})$

Boosted regression tree

Algorithm

Can be any model

load y, x, x_{new}

set parameters: d, ntrees, λ

set $\hat{f}(x_{\text{new}})$ = 0

set $r \leftarrow y$ (residuals equal to the data)

for b in 1 to ntrees

  train d split tree model on r and x

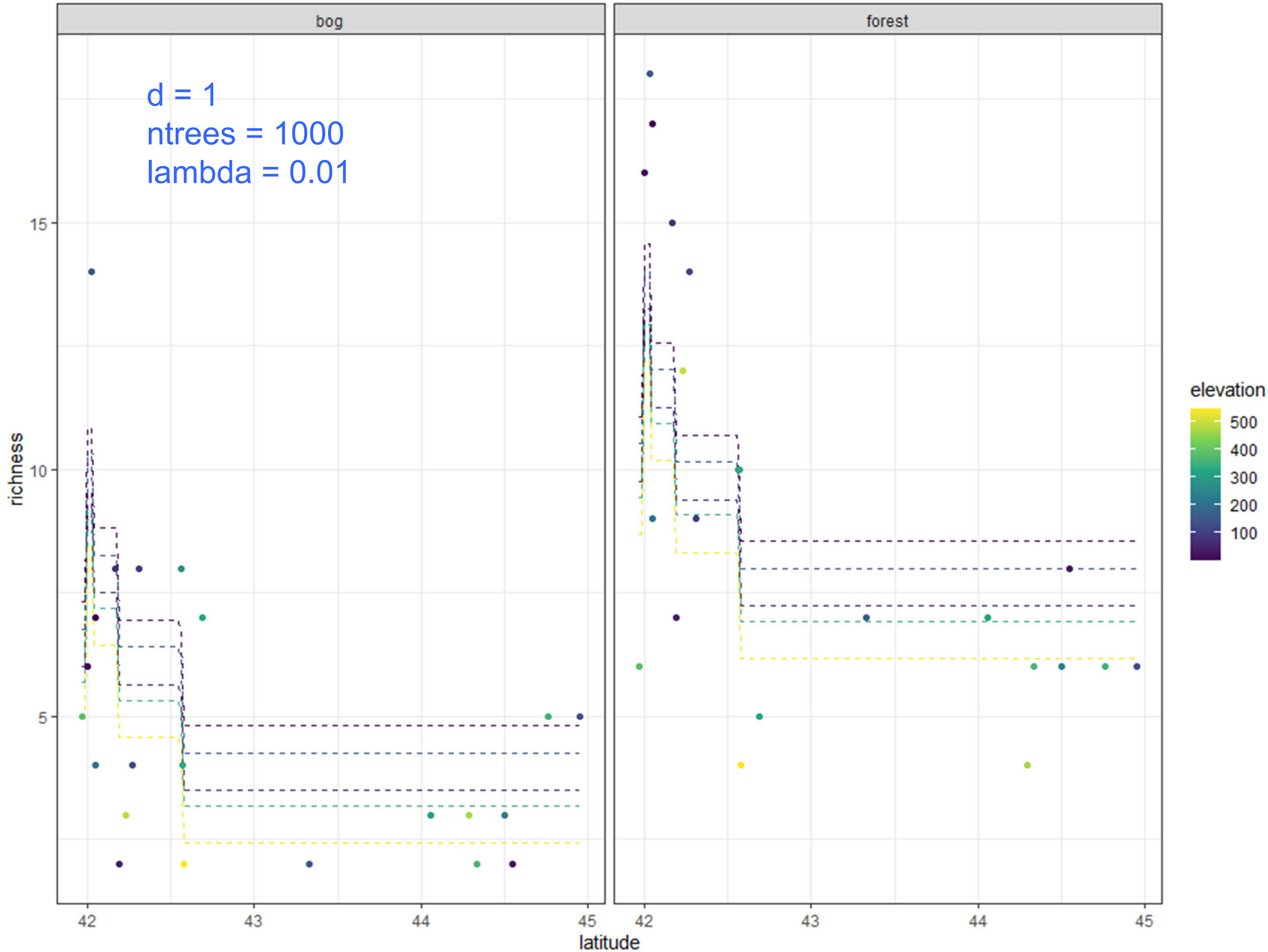
  predict residuals, $\hat{r}_b(x)$, from trained tree

  update residuals: $r \leftarrow r - \lambda\hat{r}_b(x)$

  predict y increment, $\hat{f}_b(x_{\text{new}})$, from trained tree

  update prediction: $\hat{f}(x_{\text{new}}) \leftarrow \hat{f}(x_{\text{new}}) + \lambda\hat{f}_b(x_{\text{new}})$

return $\hat{f}(x_{\text{new}})$


d = 1
ntrees = 1000
lambda = 0.01
bog
forest
richness
latitude
15
10
5
42
43
44
45
elevation
500
400
300
200
100

Boosted regression tree

Algorithm

load y, x, x_{new}

set parameters: d, ntrees, λ

set $\hat{f}(x_{\text{new}})$ = 0

set $r \leftarrow y$ (residuals equal to the data)

for b in 1 to ntrees

    train d split tree model on r and x

    predict residuals, $\hat{r}_b(x)$, from trained tree

    update residuals: $r \leftarrow r - \lambda \hat{r}_b(x)$

    predict y increment, $\hat{f}_b(x_{\text{new}})$, from trained tree

    update prediction: $\hat{f}(x_{\text{new}}) \leftarrow \hat{f}(x_{\text{new}}) + \lambda \hat{f}_b(x_{\text{new}})$

return $\hat{f}(x_{\text{new}})$

Gradient descent

Gradient descent

predict residuals, $\hat{r}_b(x)$, from trained tree
update residuals: $r \leftarrow r - \lambda \hat{r}_b(x)$

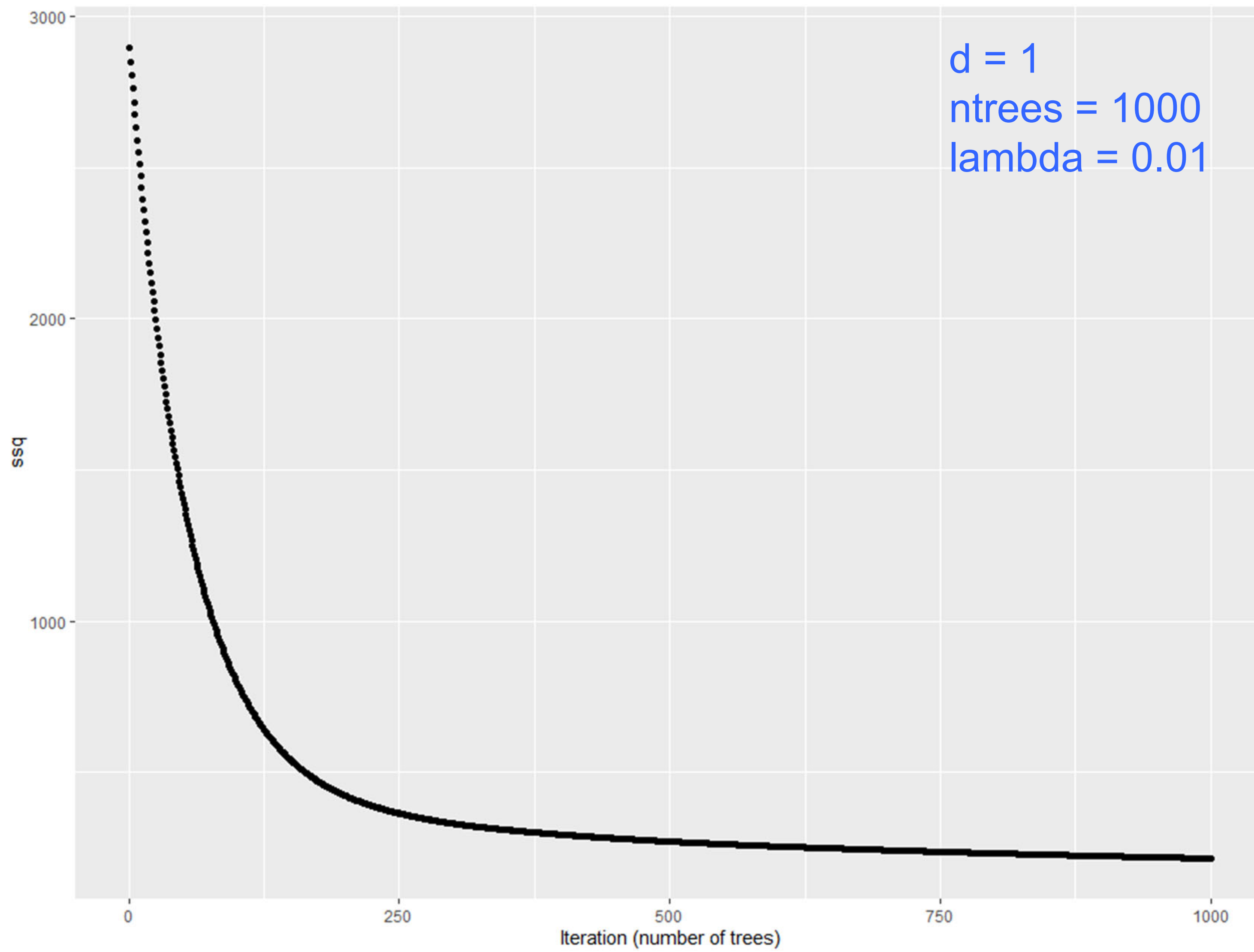
Loss function is SSQ and we are descending it's surface

Estimated gradient at x is the predicted residual $\hat{r}_b(x)$

λ is the increment taken down the gradient

r gets closer to 0 at each step, so SSQ goes down

Gradient descent



Stochastic gradient descent

Algorithm

for b in 1 to ntrees

  randomly sample rows from (r, x) ⟵ new step

  train d split tree model on the random sample

  predict residuals, $\hat{r}_b(x)$, from trained tree

  update residuals: $r \leftarrow r - \lambda \hat{r}_b(x)$

  predict y increment, $\hat{f}_b(x_{\text{new}})$, from trained tree

  update prediction: $\hat{f}(x_{\text{new}}) \leftarrow \hat{f}(x_{\text{new}}) + \lambda \hat{f}_b(x_{\text{new}})$

return $\hat{f}(x_{\text{new}})$