CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

Outline for March 6

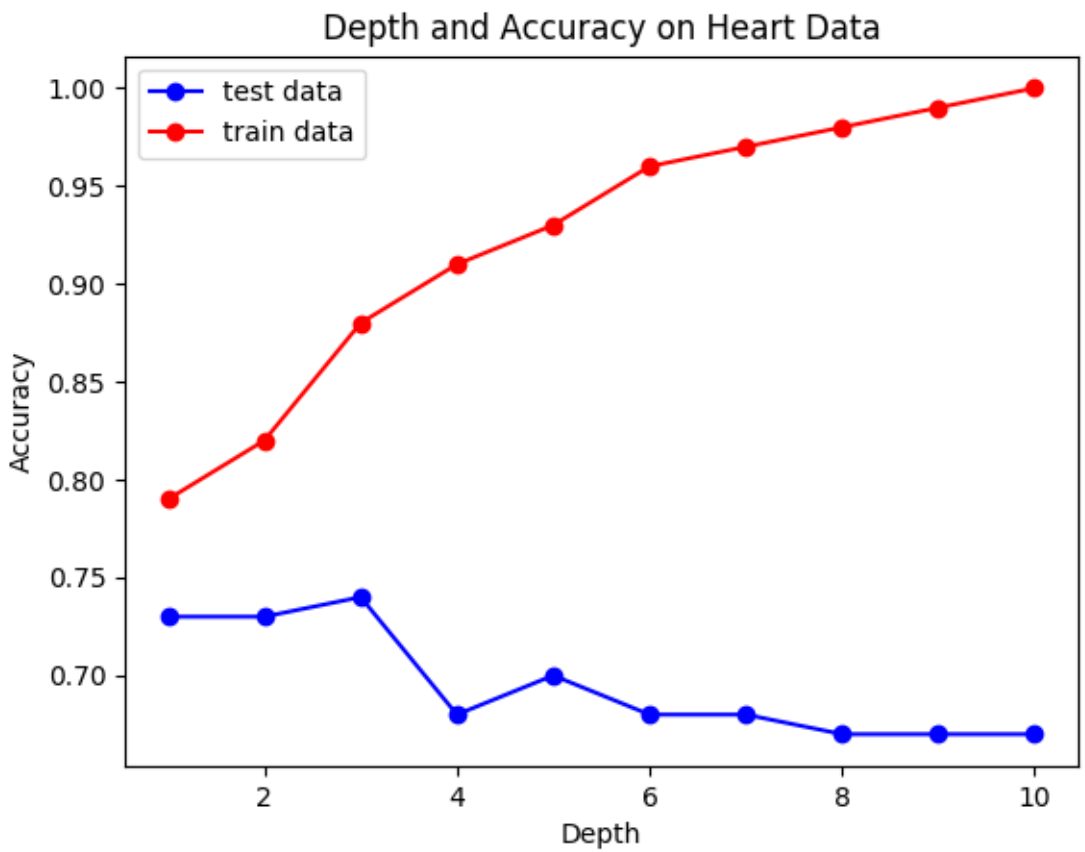
- Lab 2 examples
- Ensemble methods
  - Bagging
  - Random Forests
  - Boosting

- **Lab 4 due Friday**
- Check-in today during lab
  - should be done with one of Logistic Regression or Naïve Bayes

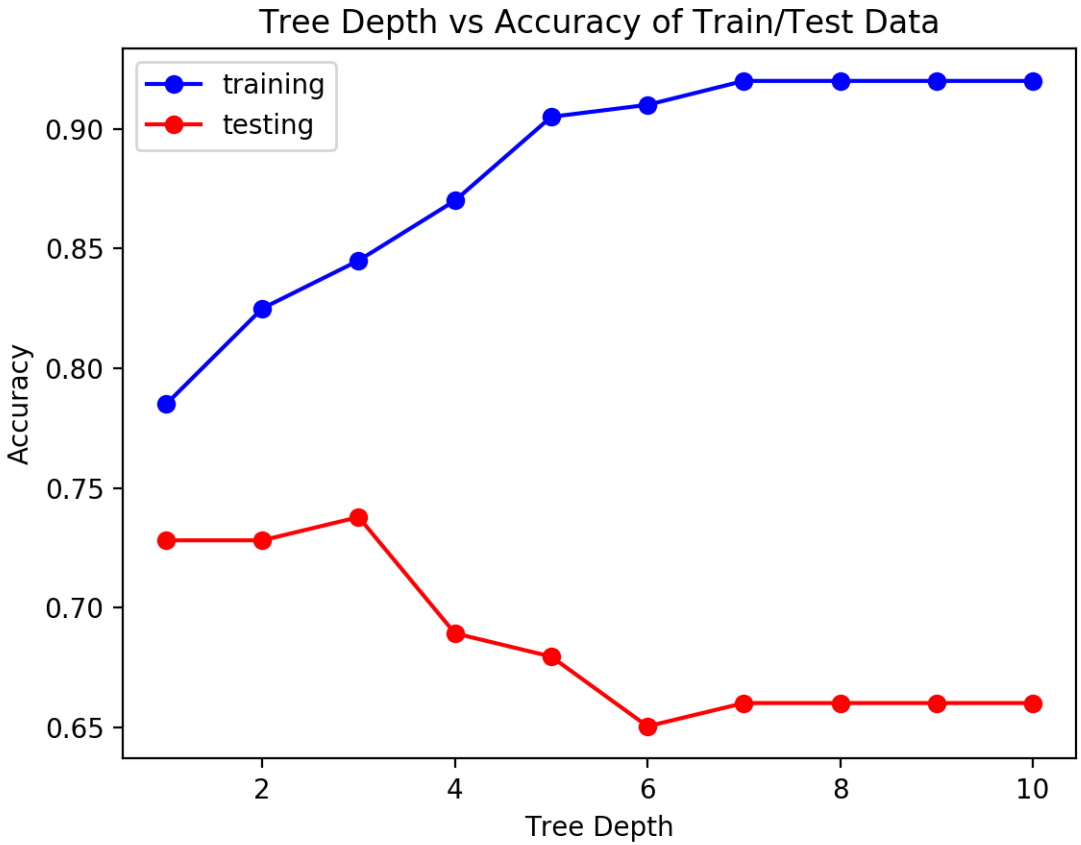
Outline for March 6

- Lab 2 examples
- Ensemble methods
  - Bagging
  - Random Forests
  - Boosting

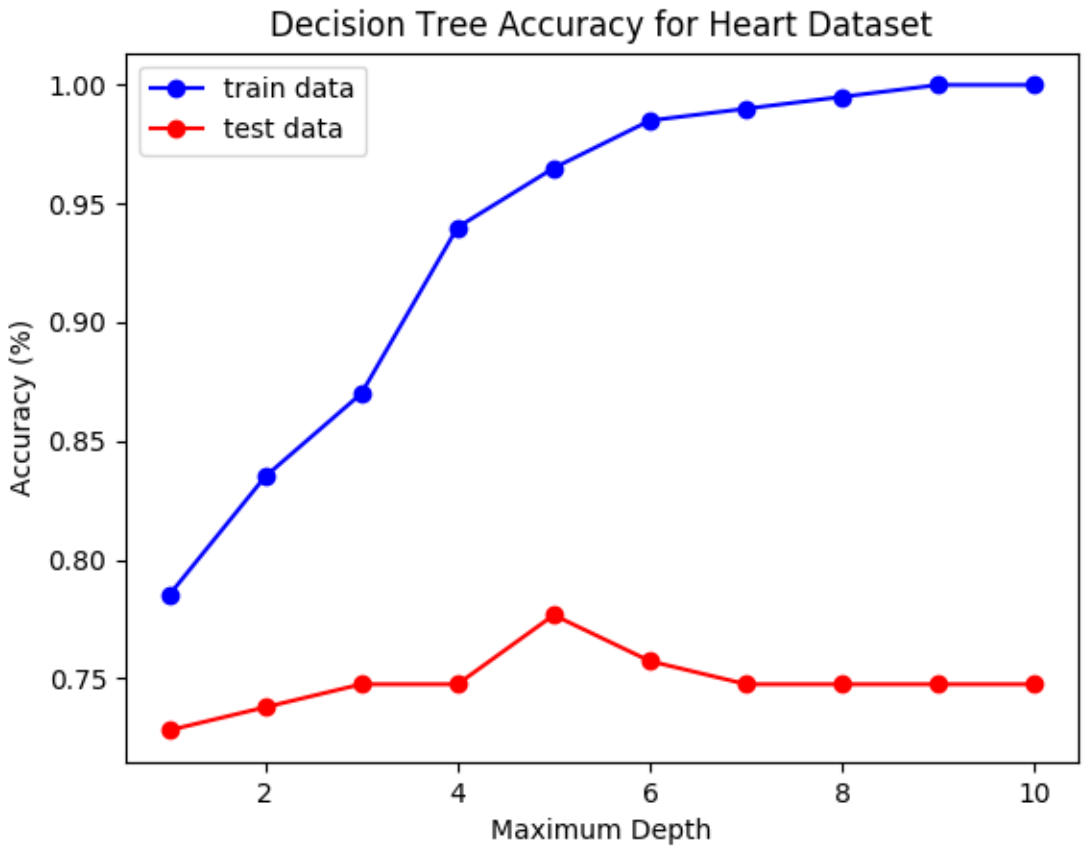
Lab 2 (heart): Henrik & Prav



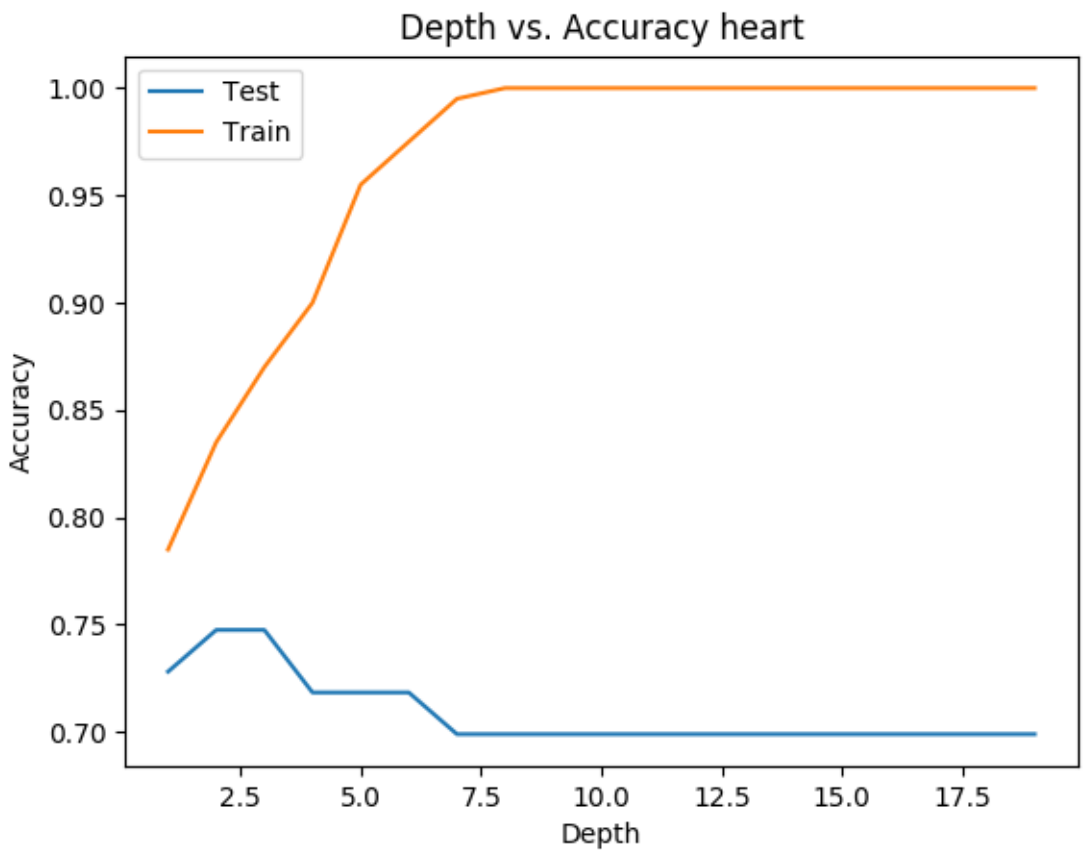
Lab 2 (heart): Lyla & Sam



Lab 2 (heart): Raymond & Kenny



Lab 2 (heart): Mikey & Dylan



Lab 2: Josh & Matthieu

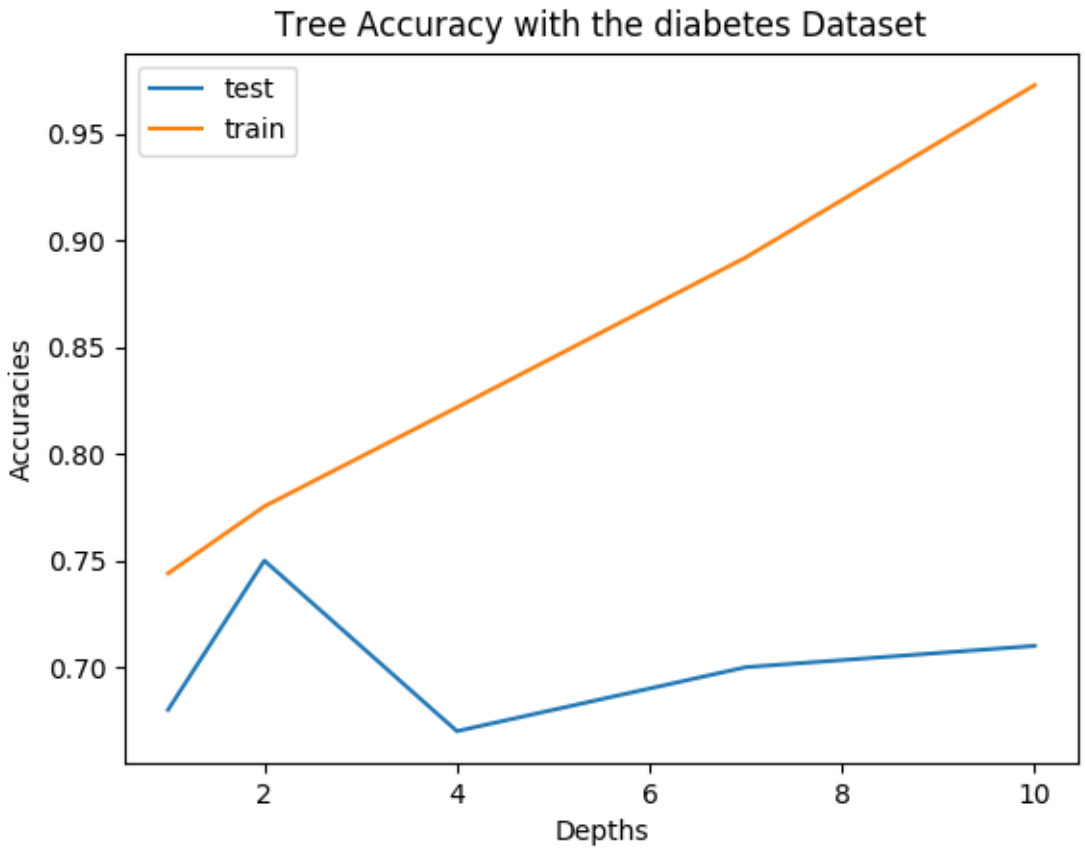
Heart

Depth	Train Accuracy	Test Accuracy
1	0.825	0.7281553398058253
2	0.86	0.7281553398058253
3	0.875	0.7184466019417476
4	0.91	0.6893203883495146
5	0.915	0.6893203883495146
6	0.91	0.6796116504854369
7	0.92	0.6601941747572816
8	0.92	0.6601941747572816
9	0.92	0.6601941747572816
10	0.92	0.6601941747572816
11	0.92	0.6601941747572816
12	0.92	0.6601941747572816
13	0.92	0.6601941747572816
14	0.92	0.6601941747572816
15	0.92	0.6601941747572816
16	0.92	0.6601941747572816
17	0.92	0.6601941747572816
18	0.92	0.6601941747572816
19	0.92	0.6601941747572816
20	0.92	0.6601941747572816

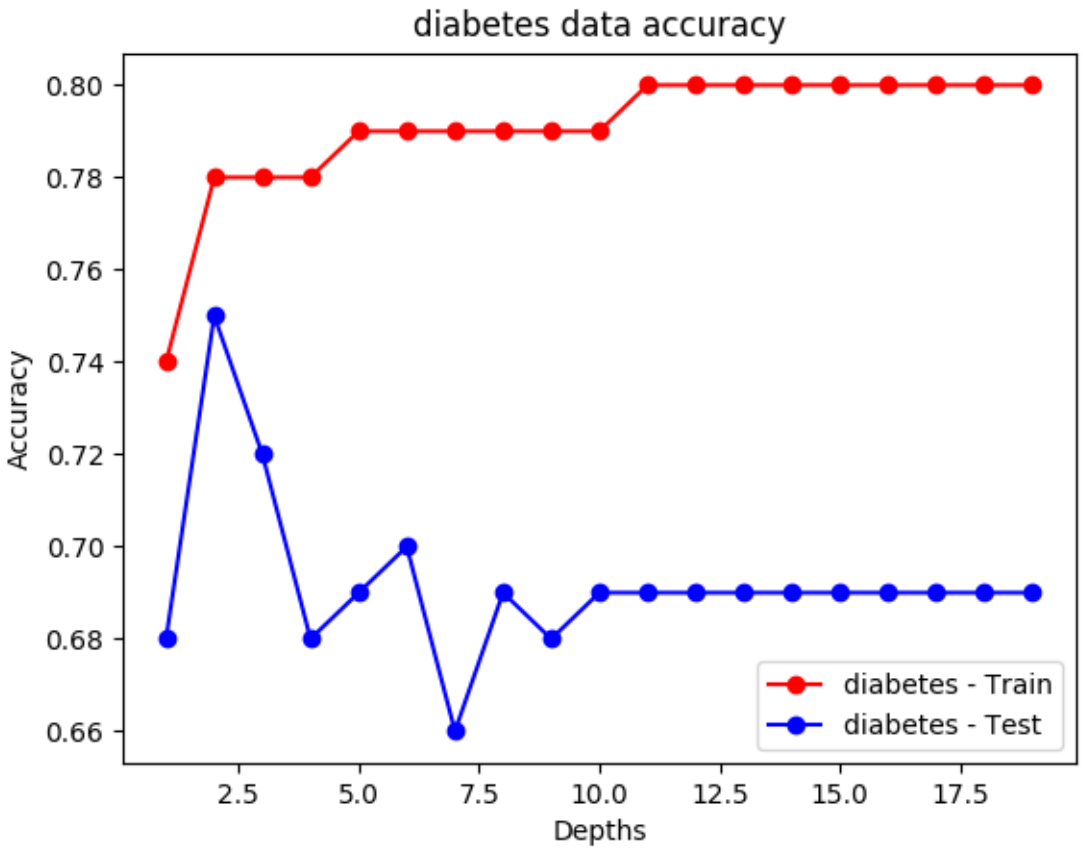
Diabetes

Depth	Train Accuracy	Test Accuracy
1	0.7544910179640718	0.7
2	0.7544910179640718	0.7
3	0.7544910179640718	0.69
4	0.7529940119760479	0.73
5	0.7964071856287425	0.69
6	0.7934131736526946	0.73
7	0.8143712574850299	0.69
8	0.8203592814371258	0.64
9	0.8218562874251497	0.65
10	0.8353293413173652	0.7
11	0.842814371257485	0.7
12	0.8458083832335329	0.69
13	0.8502994011976048	0.67
14	0.8502994011976048	0.67
15	0.8502994011976048	0.67
16	0.8502994011976048	0.67
17	0.8502994011976048	0.67
18	0.8502994011976048	0.67
19	0.8502994011976048	0.67
20	0.8502994011976048	0.67

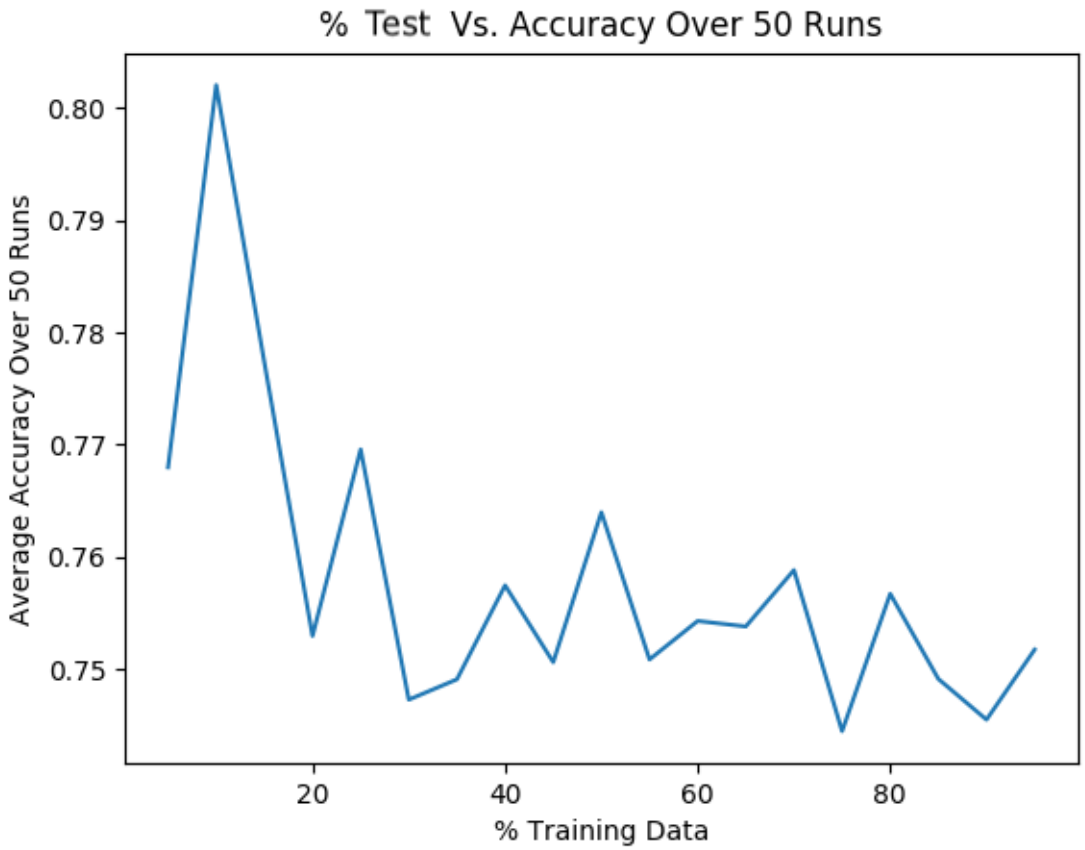
Lab 2 (diabetes): Haochen & Nav



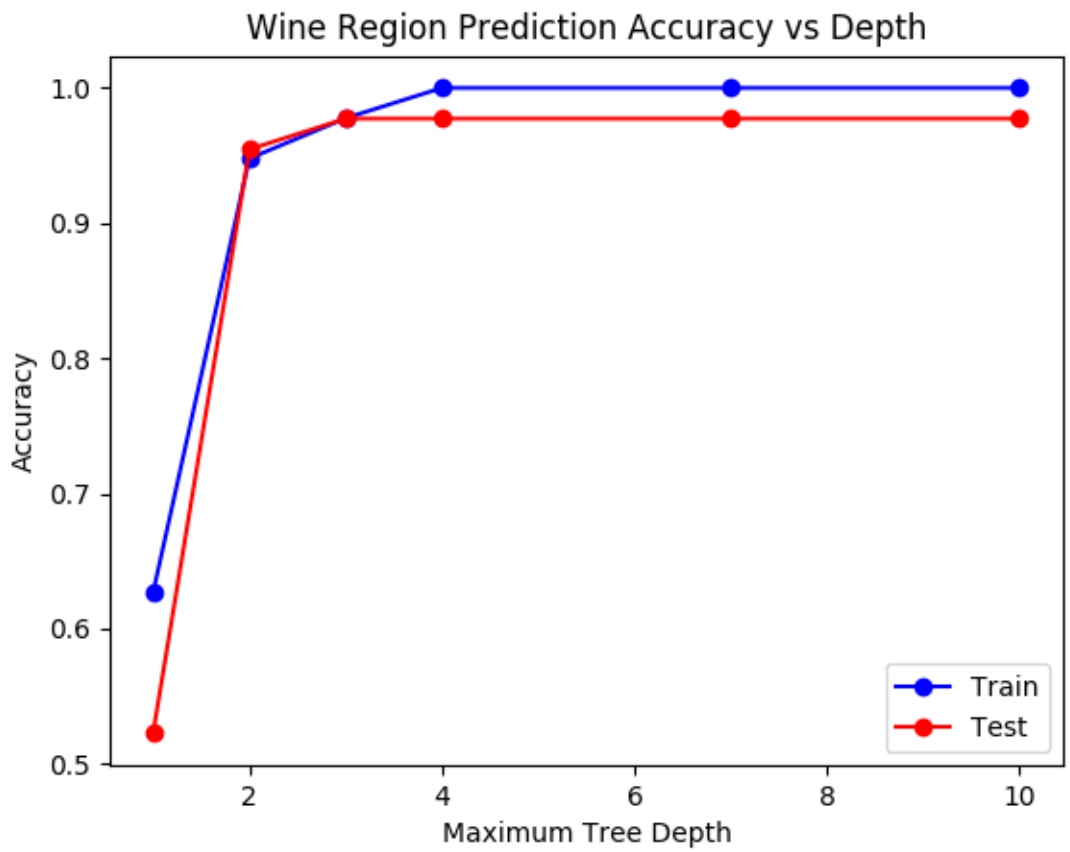
Lab 2 (diabetes): Greg & Hari



Lab 2 (percent train): Mikey & Dylan



Lab 2 (multi-class): Gabriel & Keton



Outline for March 6

- Lab 2 examples
- Ensemble methods
  - Bagging
  - Random Forests
  - Boosting

Ensemble Notation

T: # of models/classifiers (index t)

$\vec{x}$: test example

y: test label (binary)

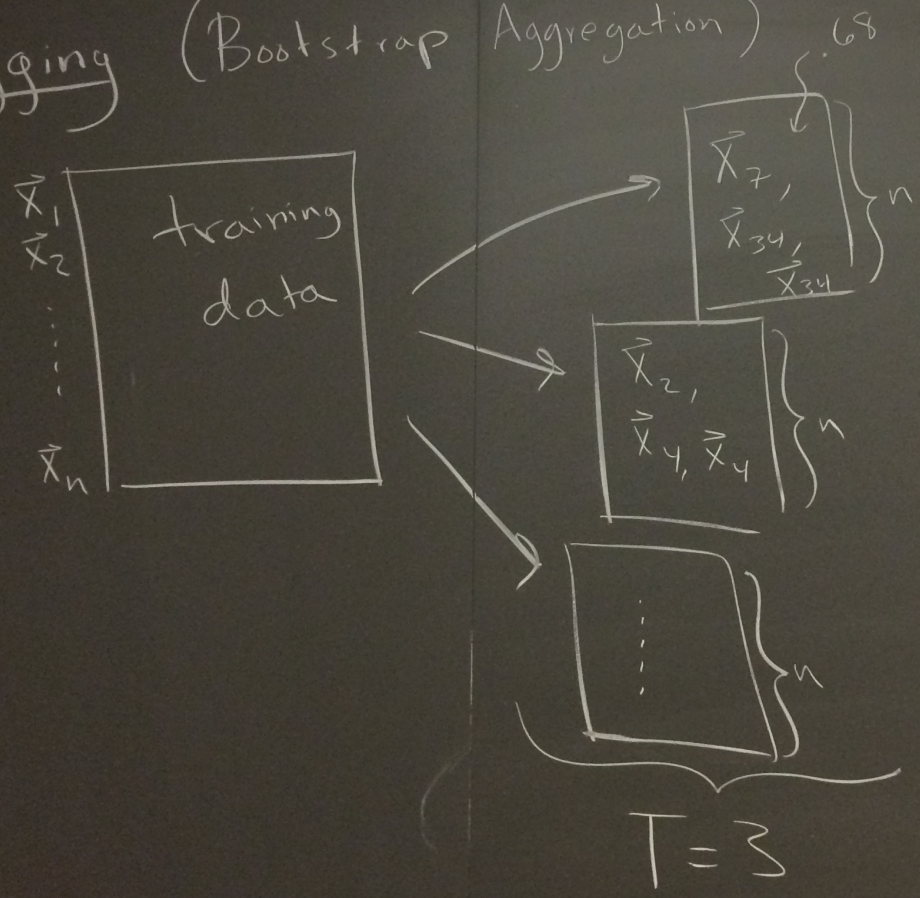
X_t: bootstrap training dataset t

$h^{(t)}(x)$: hypothesis about x from model t

r: prob of error for each model

R: # votes for wrong class

Bagging (Bootstrap Aggregation)



Why n examples?

prob didn't choose a data point:

$$\left(\frac{n-1}{n}\right)^n$$

$$= \left(1 - \frac{1}{n}\right)^n \approx e^{-1} \approx .3$$

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

$$= e^x$$

Algorithm Idea

Train for $t = 1, 2, \dots T$:

- create bootstrap sample X_t
- train on X_t to get model $h^{(t)}$

Test: for $\vec{x}$ in test:

$$h(\vec{x}) = \underset{y \in \{0,1\}}{\operatorname{argmax}} \sum_{t=1}^{T} \mathbb{1}\left(h^{(t)}(x) = y\right)$$

$$R = \sum_{t=1}^{T} \mathbb{1}\left(h^{(t)}(x) = \tilde{y}\right)$$

wrong class

$$P(R=k) = \binom{T}{k} \underbrace{r^k}_{\text{wrong } k \text{ times}} \underbrace{(1-r)^{T-k}}_{\text{right } (T-k) \text{ times}}$$

binomial distribution

"T choose k"

$$\frac{T!}{k!(T-k)!}$$

$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
0	0	1
0	1	0
1	0	0

$r^2(1-r)$

$\binom{3}{2} = 3$

$P(\text{overa}$

$= P($

$= \sum_{k=}$

$P(\text{error})$

$$R = \sum_{t=1}^{T} \mathbb{1}\left(h^{(t)}(x) = \tilde{y}\right)$$

wrong class.

$$P(R=k) = \binom{T}{k} \cdot \underbrace{r^{k}}_{\text{wrong } k \text{ times}} \underbrace{(1-r)^{T-k}}_{\text{right } (T-k) \text{ times}}$$

"T choose k"

$$\frac{T!}{k!\,(T-k)!}$$

binomial distribution

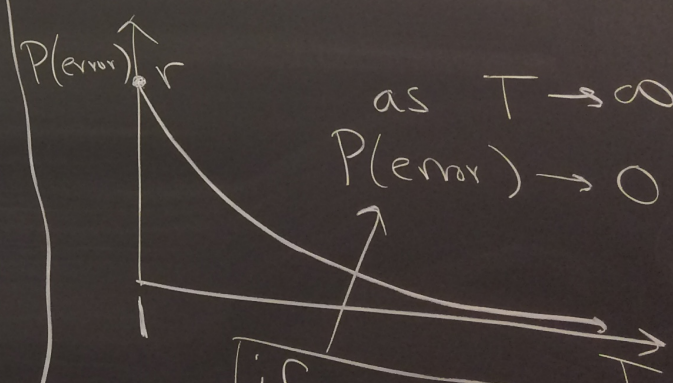
$$\sum_{t=1}^{T} \mathbb{1}\left(h^{(t)}(x) = y\right)$$

$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
0	0	1
0	1	0
1	0	0

$r^2(1-r)$

$$\binom{3}{2} = 3$$

$$P(\text{overall wrong}) = P\left(R > \frac{T}{2}\right) = \sum_{k=\frac{T+1}{2}}^{T} \binom{T}{k} r^{k} (1-r)^{T-k}$$



as $T \to \infty$

$P(\text{error}) \to 0$

$\boxed{\text{if } r < 0.5}$

Example $T = 3$

$r = \frac{1}{4}$

$\vec{x}_{test} = \boxed{}$

$y_{test} = 1$

options

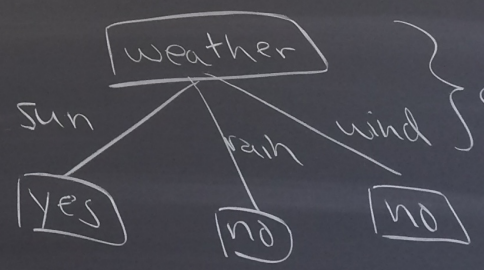
$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$P(\text{wrong})$
0	0	0	$\left(\frac{1}{4}\right)^3$
0	0	1	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
0	1	0	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
0	1	1	
1	0	0	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
1	0	1	
1	1	0	
1	1	1	

$$P(\text{error}) = \left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)\cdot 3$$

$$= \frac{10}{64} \approx \boxed{0.16}$$

compare to r and it's better!

decision stump



depth 1

1 feature

Rand

idea

goo

in

Outline for March 6

- Lab 2 examples
- Ensemble methods
  - Bagging
  - Random Forests
  - Boosting

Random Forests

idea: choose a different subset of features for every classifier t.

goal: decorrelate ~~trees~~ models

in practice: choose $\sqrt{p}$ features

* without replacement for each model
* every model independent data points and independent features

Heart data: bagging vs. random forest

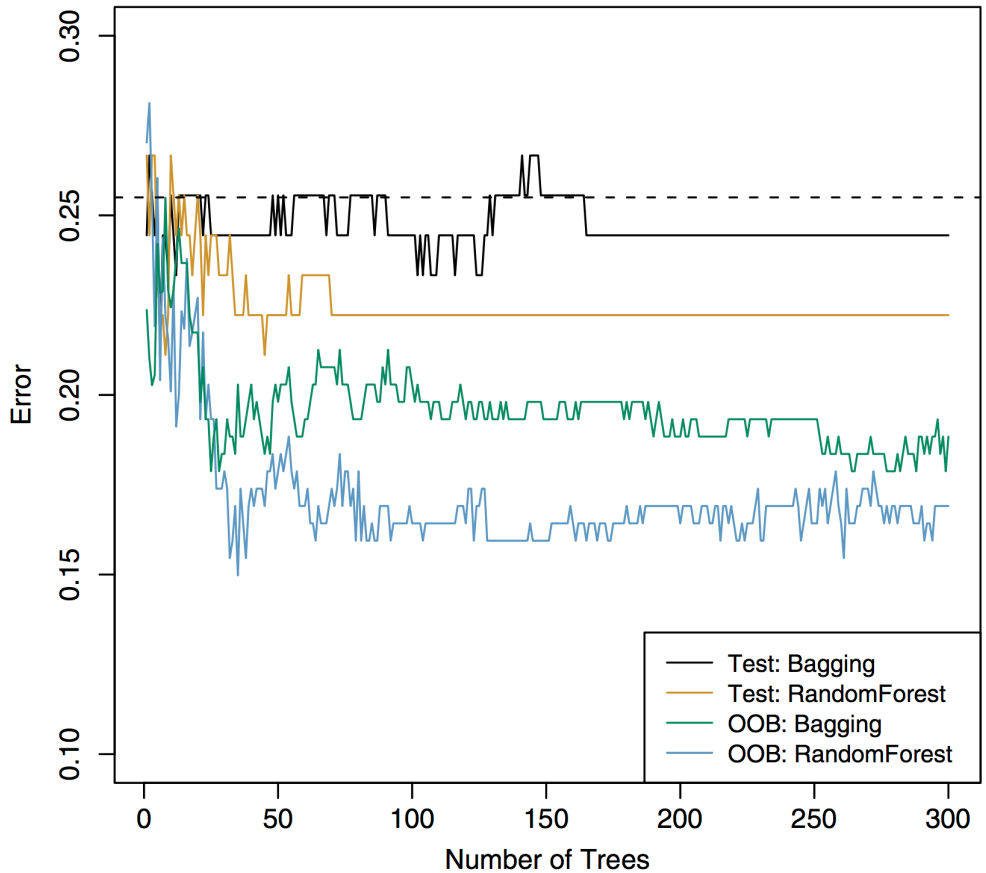


FIGURE 8.8. *Bagging and random forest results for the* `Heart` *data. The test error (black and orange) is shown as a function of B, the number of bootstrapped training sets used. Random forests were applied with $m = \sqrt{p}$. The dashed line indicates the test error resulting from a single classification tree. The green and blue traces show the OOB error, which in this case is considerably lower.*

ISL: Fig 8.8

Heart data: most important features

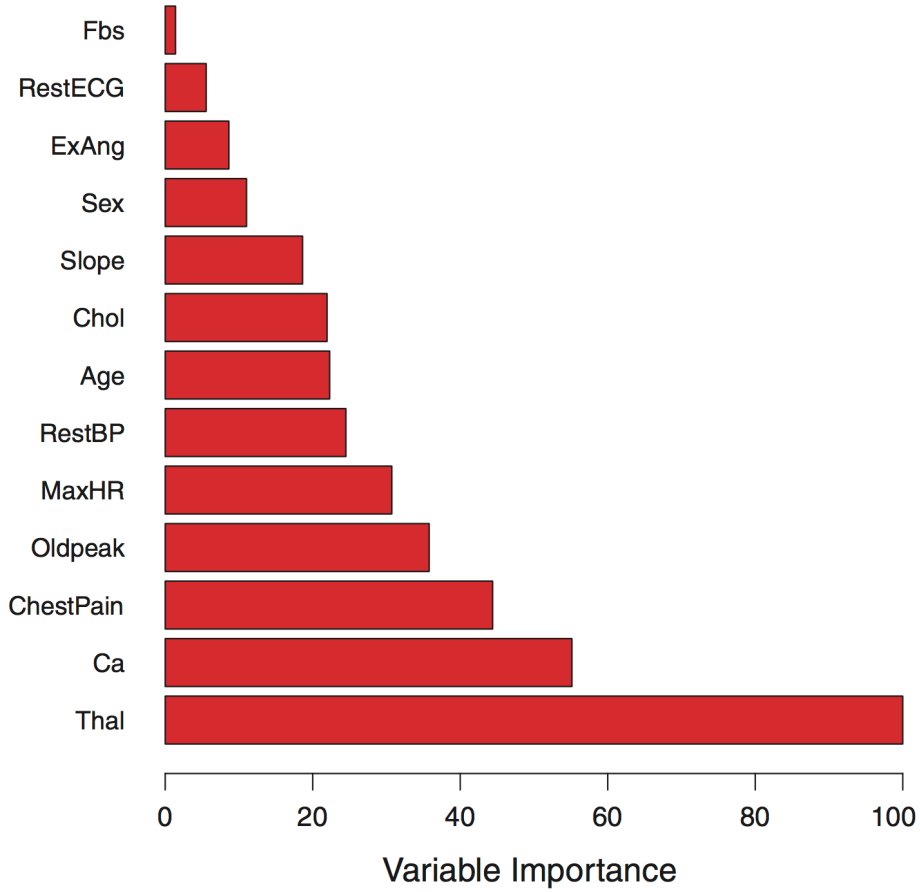


FIGURE 8.9. *A variable importance plot for the* `Heart` *data. Variable importance is computed using the mean decrease in Gini index, and expressed relative to the maximum.*

ISL: Fig 8.9