CS 383: Machine Learning

Prof Adam Poliak

Fall 2024

10/30/2024

Lecture 18

Announcements

Thursday reading quiz: Duame Textbook Chapter 13

Outline

Ensemble Methods

- Bagging
- Boosting
- Weighted Entropy

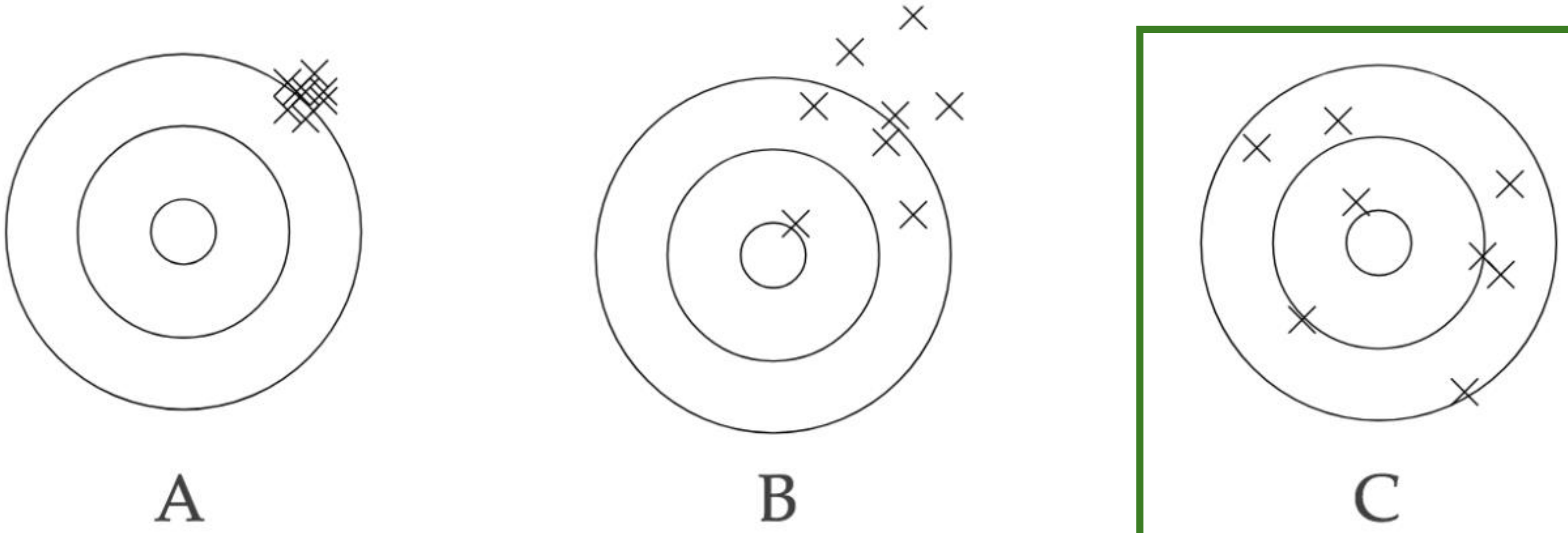
Outline

Logistic Regression

Ensemble Methods

- Bagging
- Boosting
- Weighted Entropy

Quiz: recap bias and variance



	A		B		C
Variance:	low	Variance:	high	Variance:	high
Bias:	high	Bias:	high	Bias:	low

This is the type of classifier we want to average!

Label each picture with variance (high or low) and bias (high or low)

Example from Ameet Soni

Ensemble Intuition

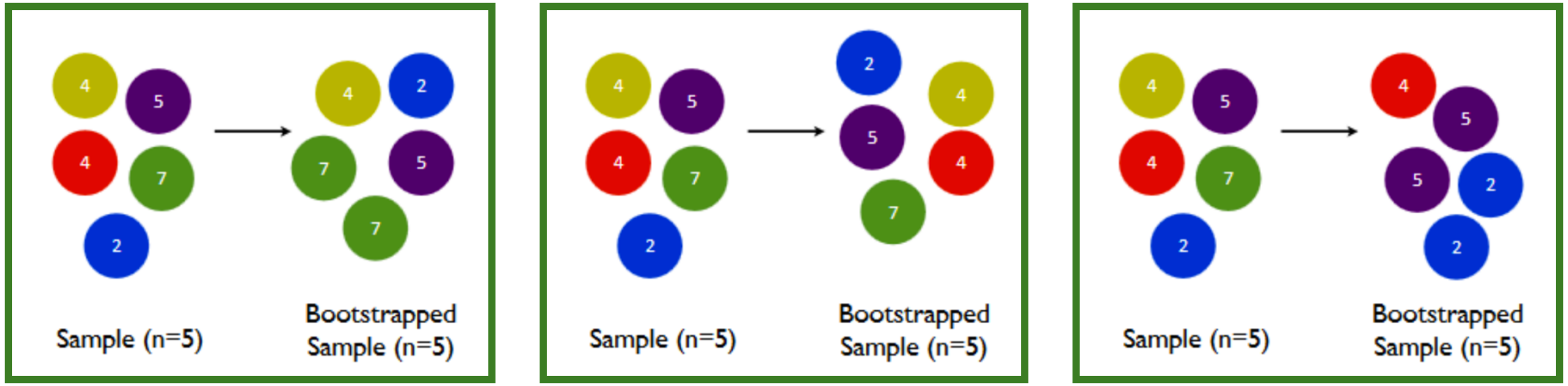
Average the results from several models with high variance and low bias

- Important that models be diverse (don't want them to be wrong in the same ways)

If n observations each have variance s^2, then the mean of the observations has variance s^2/n (reduce variance by averaging!)

Bagging Algorithm

- Bagging = Bootstrap Aggregation [Brieman, 1996]
- *Bootstrap* (randomly sample <u>with replacement</u>) original data to create many different training sets
- Run base learning algorithm on each new data set independently



Desmond Ong, Stanford

Modified from Ameet Soni

Notation

T: # of models/classifiers

x: test example

$X^{(t)}$: bootstrap training set t

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

r: probability of error of individual model

R: number of votes for wrong class

Bagging Algorithm

Train

Generate $X^{(t)}$ for $t = 1, \ldots, T$ using bootstrap sampling

Train classifier $h^{(t)}$ on $X^{(t)}$

Test

for x in test data:

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

Probability that $R = k$?

$$P(R = k) = \binom{T}{k} r^k (1 - r)^{T-k}$$

What is probability that ensemble is wrong?

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

If $r < \frac{1}{2}, \lim_{T \to \infty} P\left(R > \frac{T}{2}\right) = 0$

Random Forest

Idea: choose a different subset of features for every classifier t

Choose weak/base classifiers

Typically use *decision stumps* (depth 1)

Goal: decorrelate models

In practice: choose sqrt(p) features

- Without replacement for each model
- Every model: data points and features chosen independently

Outline

Logistic Regression

Ensemble Methods

- Bagging
- **Boosting**
- Weighted Entropy

Boosting Algorithm

Train

Assign equal weights to all training examples $(\frac{1}{n})$

For T iterations:

- Learn classifier using weighted examples
- Change example weights based on training error

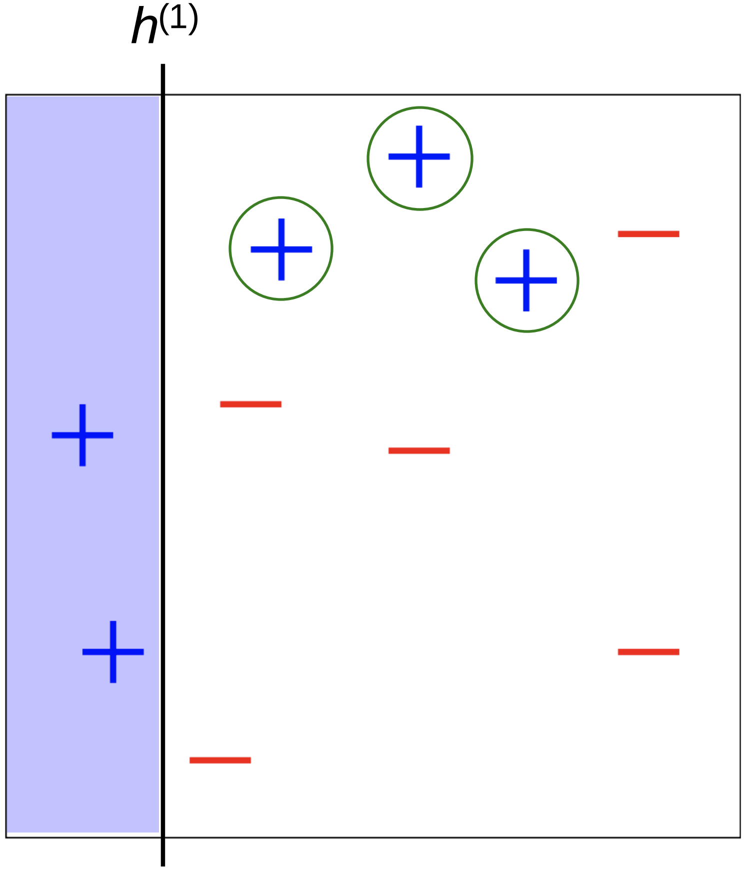
Test

Get predictions from T classifiers

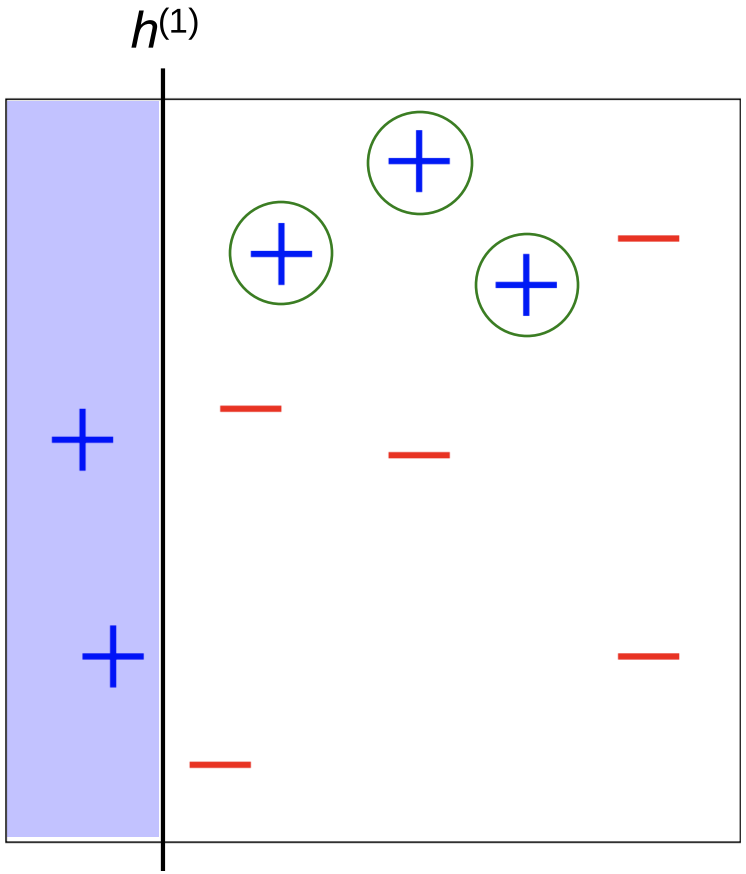
Weighted vote among T classifiers

- weight is based on how well classifier t performed during training

Handout: Round 1



Handout: Round 1



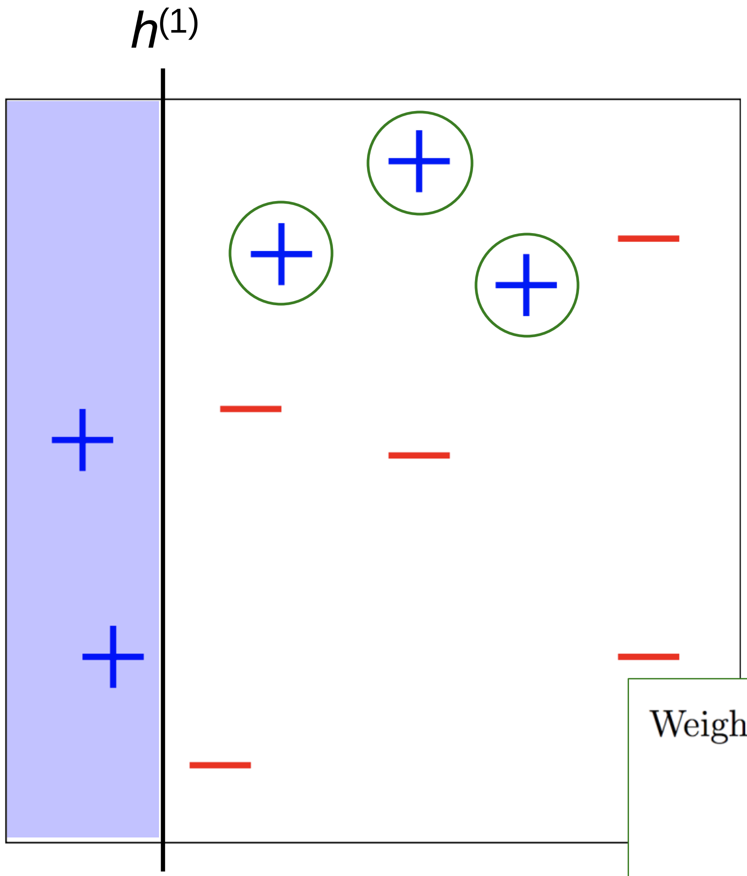
$w_i^{(1)} = \frac{1}{10}$ for all $i = 1, 2, \cdots, 10.$

$\epsilon_1 = \frac{3}{10}$ (three points incorrectly classified, all with weight $\frac{1}{10}$)

$\alpha_1 = \frac{1}{2} \ln\left(\frac{1-\frac{3}{10}}{\frac{3}{10}}\right) = \ln\sqrt{\frac{7}{3}} \approx 0.42$

- correctly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp\left(-\ln\sqrt{\frac{7}{3}}\right)$
- incorrectly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp\left(\ln\sqrt{\frac{7}{3}}\right)$

Handout: Round 1



$w_i^{(1)} = \frac{1}{10}$ for all $i = 1, 2, \cdots, 10.$

$\epsilon_1 = \frac{3}{10}$ (three points incorrectly classified, all with weight $\frac{1}{10}$)

$\alpha_1 = \frac{1}{2} \ln \left(\frac{1 - \frac{3}{10}}{\frac{3}{10}} \right) = \ln \sqrt{\frac{7}{3}} \approx 0.42$

- correctly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left(-\ln \sqrt{\frac{7}{3}} \right)$
- incorrectly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left(\ln \sqrt{\frac{7}{3}} \right)$

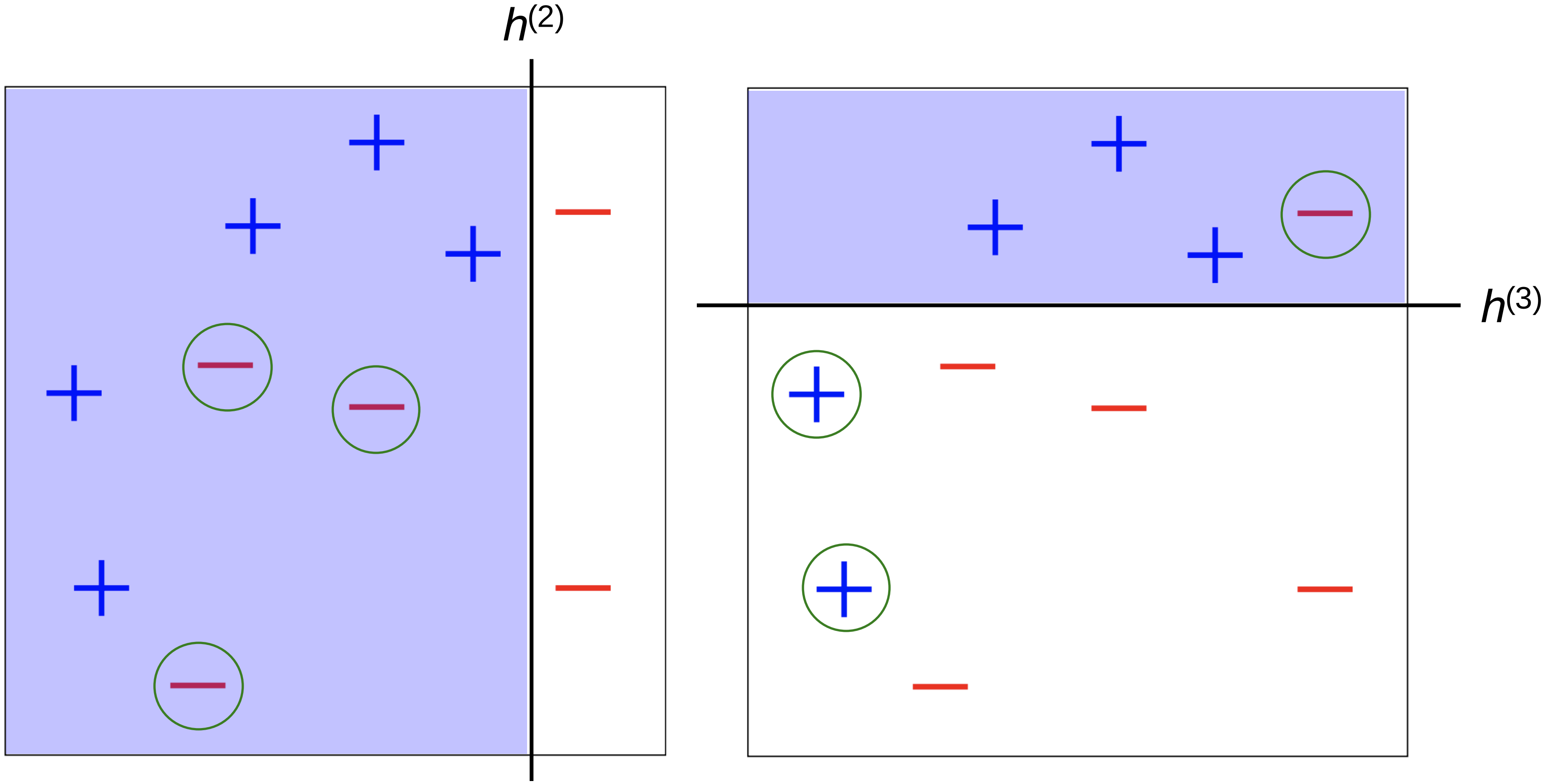
Weights must sum to 1, $\Rightarrow$

$$7 \cdot \frac{c_1}{10} \exp \left(-\ln \sqrt{\frac{7}{3}} \right) + 3 \cdot c_1 \cdot \frac{1}{10} \exp \left(\ln \sqrt{\frac{7}{3}} \right) = 1$$

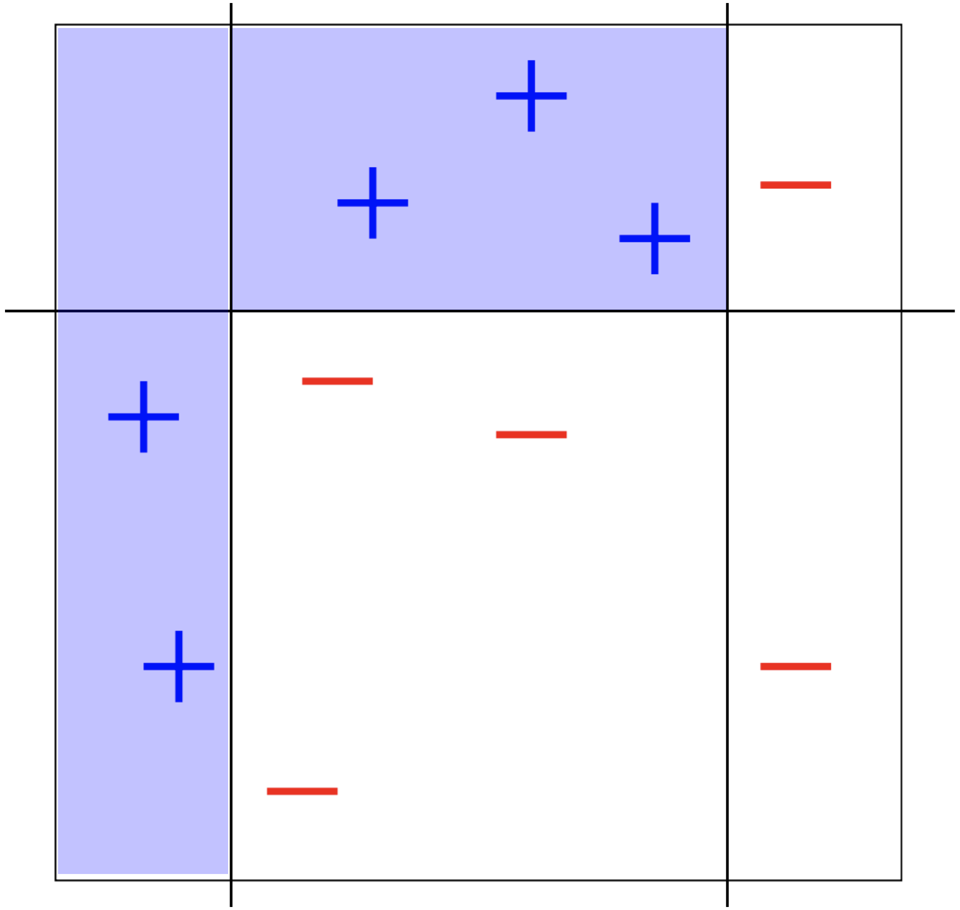
$$\Rightarrow \quad c_1 = \frac{5}{\sqrt{21}}$$

- correctly classified: $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{3}{7}} = \frac{1}{14}$ decrease!
- incorrectly classified: $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{7}{3}} = \frac{1}{6}$ increase!

Handout: Round 2 & 3 (exercise!)



Handout: final classifier



$$h(\boldsymbol{x}) = \text{sign}\left(0.42 \cdot h^{(1)}(\boldsymbol{x}) + 0.65 \cdot h^{(2)}(\boldsymbol{x}) + 0.92 \cdot h^{(3)}(\boldsymbol{x})\right)$$