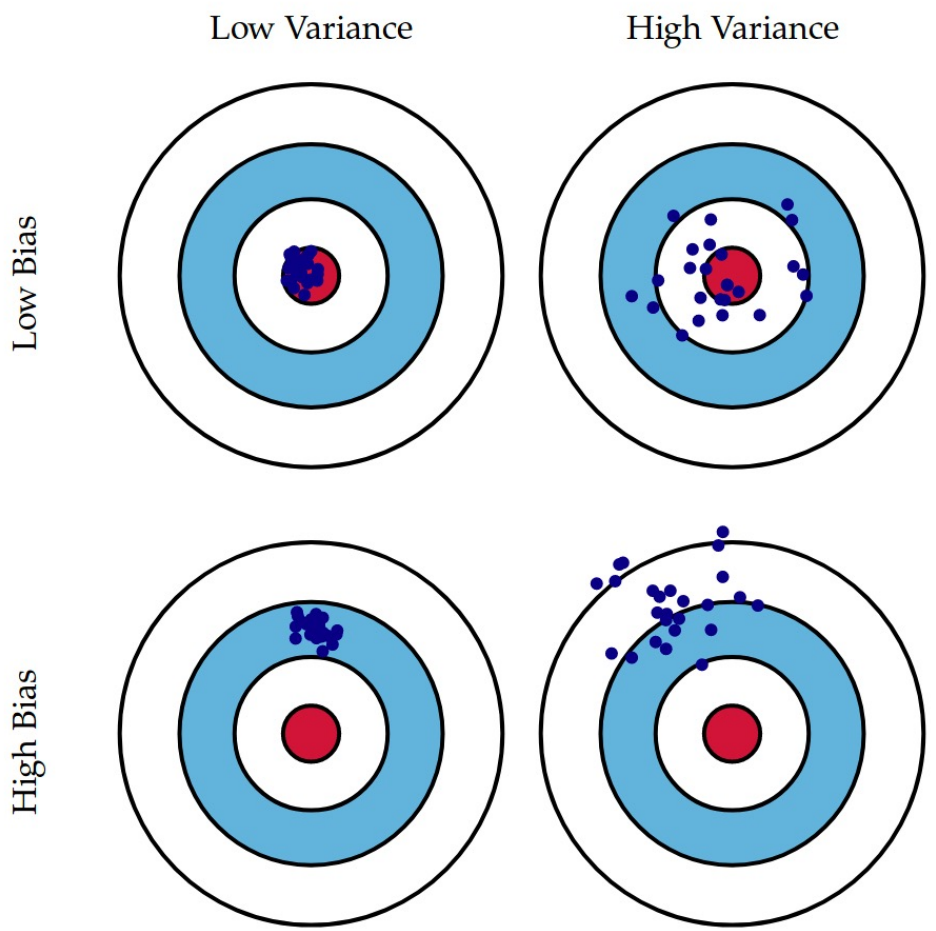
Ensembles

Exercise 2

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

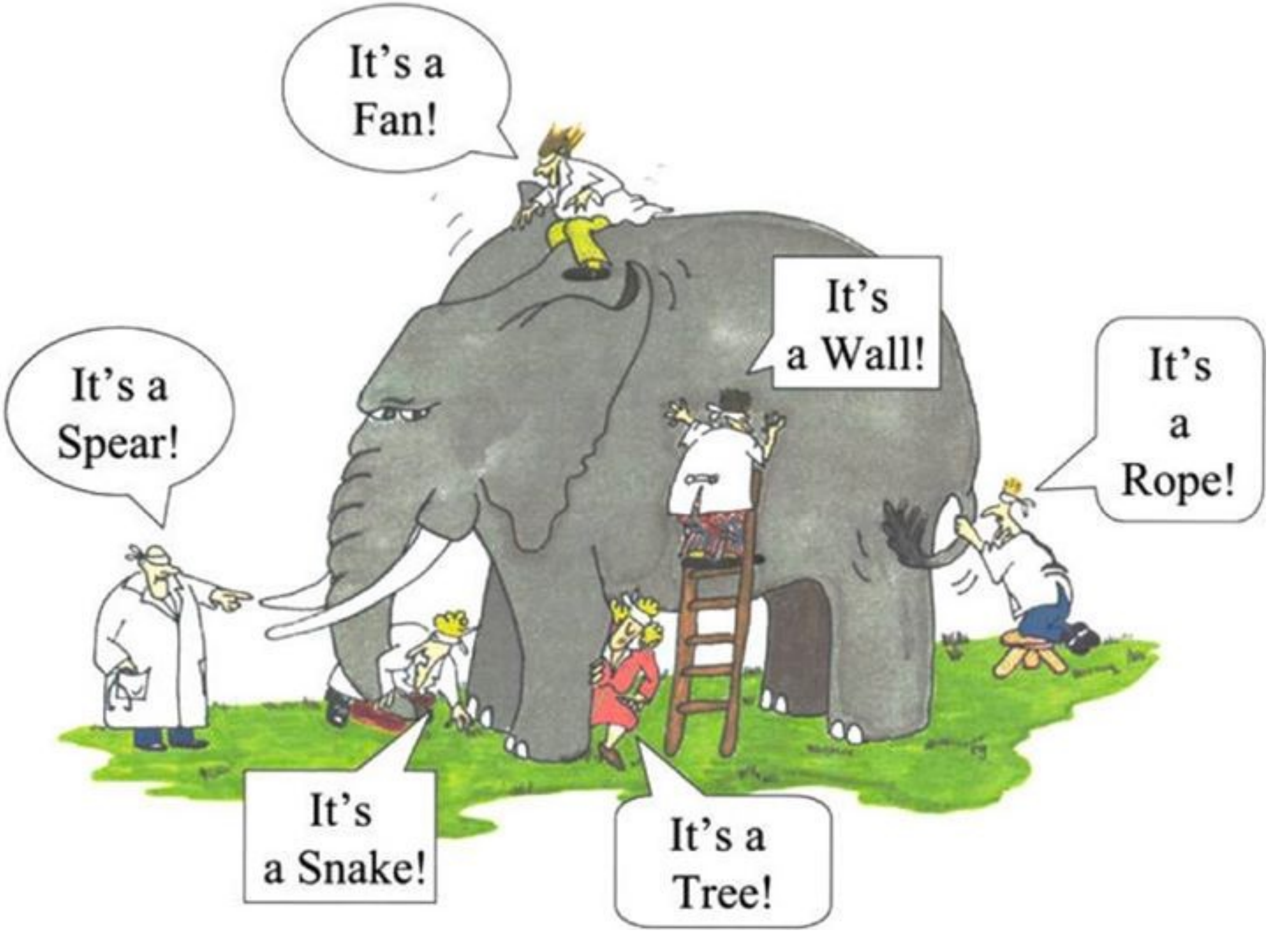
Variance versus Bias



Source: https://towardsdatascience.com/bias-and-variance-in-linear-models-e772546e0c30

The Idea of Ensemble Learning

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



Source: https://www.patheos.com/blogs/driventoabstraction/2018/07/blind-men-elephant-folklore-knowledge/

Ensembles

Yes or No?

With the Voting mechanism, predictions of multiple classifiers or regressors are combined (using the mode or the average) to find a final prediction.

Correct?

Ensembles
Yes or No?

Bagging is like voting, but instead of using differing classifiers or regressors, we split the dataset into *n* parts and vote over the results from the classifiers/regressors applied to the individual datasets.

Correct?

Ensembles

Yes or No?

Boosting is like voting, but instead of doing a majority vote (mode/average), we use a classifier to make the final decision.

Correct?

Ensembles

Yes or No?

The Random Forest algorithm is an application of the Stacking mechanism.

Correct?

Ensembles

Yes or No?

The MetaCost algorithm is an ensemble mechanism that tries to relabel the final predictions in such a way that the ”risk” of the predictions is minimized.

Correct?

Voting
Exercise

DATA

#	Brand	Dealer	Bundle	Capacity	Color	Bestseller
1	Apple	ABC Phones	yes	128	white	yes
2	Apple	ABC Phones	yes	128	black	yes
3	Apple	ABC Phones	no	64	silver	yes
4	Apple	ABC Phones	no	32	red	no
5	LG	Phones Unlimited	yes	32	white	no
6	LG	Phones Unlimited	no	64	silver	yes
7	Samsung	Phones Unlimited	yes	32	white	no
8	Samsung	My Phone 123	no	64	silver	yes
9	Samsung	My Phone 123	yes	128	red	no
10	Samsung	My Phone 123	no	128	blue	no

BASE CLASSIFIERS

(for the Bestseller attribute)

(A) Brand=Apple ➞ yes, else no

(B) Dealer=ABC Phones ➞ yes, else no

(C) Capacity=64 ➞ yes, else no

TASKS

(1) What is the accuracy of each base classifier?

(2) What is the accuracy of voting? Is it higher or lower than the best base classifier? Why?

Stacking

Exercise

DATA

#	Brand	Dealer	Bundle	Capacity	Color	Bestseller
1	Apple	ABC Phones	yes	128	white	yes
2	Apple	ABC Phones	yes	128	black	yes
3	Apple	ABC Phones	no	64	silver	yes
4	Apple	ABC Phones	no	32	red	no
5	LG	Phones Unlimited	yes	32	white	no
6	LG	Phones Unlimited	no	64	silver	yes
7	Samsung	Phones Unlimited	yes	32	white	no
8	Samsung	My Phone 123	no	64	silver	yes
9	Samsung	My Phone 123	yes	128	red	no
10	Samsung	My Phone 123	no	128	blue	no

CLASSIFIER ACCURACY

#	(A)	(B)	(C)	Voting	Bestseller
1	yes	yes	no	yes	yes
2	yes	yes	no	yes	yes
3	yes	yes	yes	yes	yes
4	yes	yes	no	yes	no
5	no	no	no	no	no
6	no	no	yes	no	yes
7	no	no	no	no	no
8	no	no	yes	no	yes
9	no	no	no	no	no
10	no	no	no	no	no
ACC	0.7	0.7	0.8	0.7	

TASKS

(1) Based on the classifier accuracies, could stacking help more than simple voting? If yes, what would be the final accuracy?

(2) Given the additional classifier (D) Capacity=32 ➞ no, else yes would the accuracy of the stacking classifier change?

Recap: MetaCost

- Conditional risk R(i|x) is the expected cost of predicting that x belongs to class i
  - $R(i|x) = \sum P(j|x)C(i, j)$
  - C(i,j) are the classification costs (classify an example of class j as class i)
  - P(j|x) are obtained by running the bagged classifiers
- The goal of MetaCost procedure is: to relabel the training examples with their “optimal” classes
  - i.e., those with lowest risk
- Then, re-run the classifier to build a final model
  - the resulting classifier will be defensive, i.e., make low-risk predictions
  - in the end, the costs are minimized

2/24/20 Heiko Paulheim 68

MetaCost Algorithm
HIV Test

- x_1: [-], $P([-]|x_1) = 0.95$
- x_2: [+], $P([+]|x_2) = 0.87$

HIV TEST		predicted	
		[+]	[-]
actual	[+]	0	1000
	[-]	20	0

Risk values:

- $R([-]|x_1) =$
- $R([+]|x_1) =$
- $R([-]|x_2) =$
- $R([+]|x_2) =$