Ensembles & diversity

- Brief introduction
- Factors influencing accuracy
- Single / Multiple-engine Ensembles
- Sampling Techniques
- Measuring diversity (CFD)

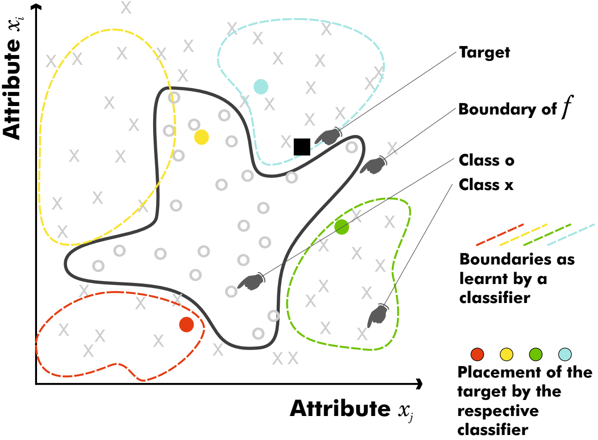
Symbol table:

h	(hypothesis) - single classifier	f	an unknown function
H	$H=\{h_1, h_2 ..., h_n\}$	x	an attribute or feature used by f to calculate an outcome y
|H|	Size of H (n in above example)		
E	Ensemble	X	collection of attributes x
S	Decision Fusion Strategy	$h[X_{training}, X_{testing}]$	
div(x)	diversity within x		h constructed with $X_{training}$ and tested with $X_{testing}$

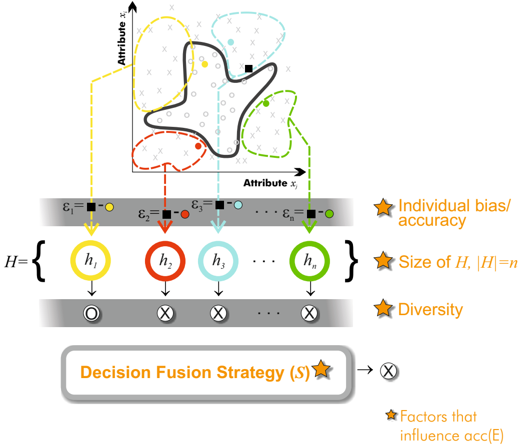
Ensembles

Ensembles

- A committee (H) of experts (h)
- Could make more accurate predictions than the individual experts
- Potentially more resilient to noise and capable of dealing with new data



2 class classification problem



Ensemble (naive representation)

Single Engine Ensembles

- $h_1;h_2;h_3;\ldots;h_n$ constructed using the same statistical learning technique.
  (If h_x is a Decision Tree then H is a Random Forest)

$$H=\left\{ \quad h_1 \quad h_2 \quad h_3 \quad \cdots \quad h_n \quad \right\}$$

Multi-engine Ensembles

- $h_1;h_2;h_3;\ldots;h_n$ constructed using different statistical learning techniques.

$$H=\left\{ h_1 \quad h_2 \quad h_3 \quad \cdots \quad h_n \right\}$$

- Does not have to be a statistical learning engine (could include a traditional Expert System or a Rules Based Decision System).
- Operates in the same problem domain, but need not respond to the same inputs.

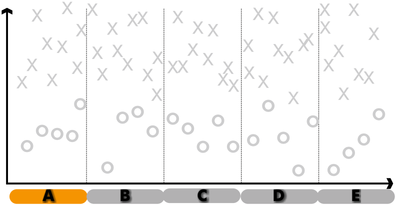
Single vs Multi-engine Ensembles

- Diversity of engine is no guarantee of diversity of results.
- Several algorithms exist to create Single Engine Ensembles, but two principle techniques:
  - Bagging
  - Boosting

Things to consider when selecting a sampling technique:

- For any given h_x: will a test case be isolated from or replaced into the training case?
- Is the training set big enough?
- Is the training set balanced?

Check the answers to these questions against the fundamental statistical assumptions of the chosen machine learning algorithm(s).



h_A : [training = X_B U X_C U X_D U X_E], [testing = X_A]

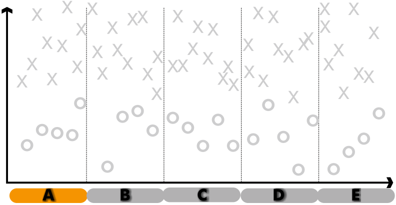
h_B : [training = X_A U X_C U X_D U X_E], [testing = X_B]

h_C : [training = X_A U X_B U X_D U X_E], [testing = X_C]

etc.

X_A refers to the collection of attributes and documents that make up subset A of the problem domain. (Not to be confused with "x" which is one of two different cases in a 2-case problem "o" & "x")

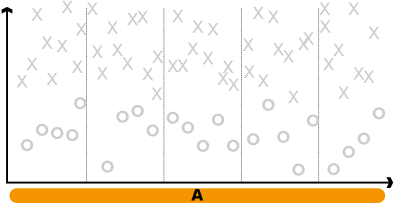
Sampling Techniques: Bagging (etc)



h_A : [training = X_B U X_C U X_D U X_E], [testing = X_A]

If $|X_A| = |X_B| = |X_C| = |X_D|$ etc
Then this is "Cross-fold Validation".

Variations include: Leave 1 out (or leave *p* out).



h_A : [training = X_A], [testing = X_A]

- ✓$_A$ correct
- ✗$_A$ incorrect

h_B : [training = {$h(X_A)$=✗}],
[testing = X_A]

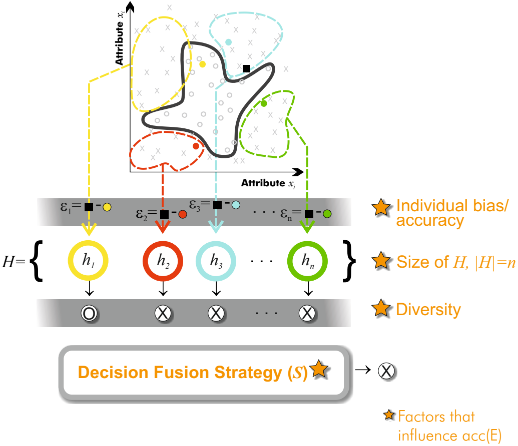
- ✓$_B$ correct
- ✗$_B$ incorrect

h_C : [training = {$h(X_B)$=✗}],
[testing = X_A]

- ✓$_C$ correct
- ✗$_C$ incorrect

Each progressive machine uses the failures from the past machines to train (or more accurately assigns a higher learning value to past mistakes).

Sampling Techniques: Boosting (etc)

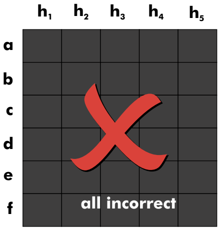


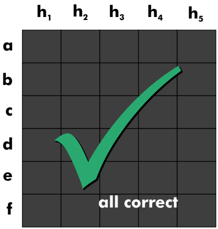
Ensemble (naive representation)

Metrics

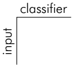
- div(H) = Coincidental Failure Diversity (CFD):
  - p_k = probability that k members of H will make the wrong choice at the same time
  - p_0 = special case - no members are wrong

$$
\mathrm{CFD} = \begin{cases} 0 & p_0 = 1 \\ 1/(1-p_0) \displaystyle\sum_{k}^{|H|} \frac{|H|-k}{|H|-1} p_k & p_0 < 1 \end{cases}
$$

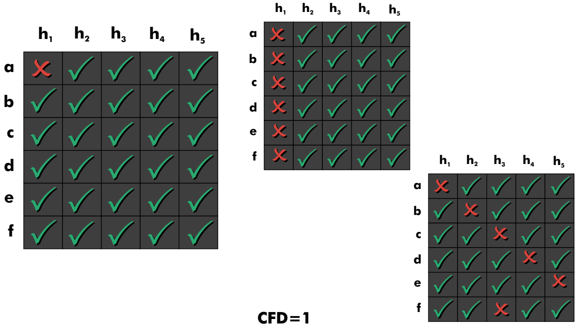




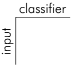
CFD=0
no diversity



Metrics: Coincidental Failure Diversity



CFD=1
maximum coincidental failure diversity



Metrics: Coincidental Failure Diversity

	h_1	h_2	h_3	h_4	h_5
a	✗	✗	✓	✓	✓
b	✓	✓	✓	✓	✓
c	✓	✓	✓	✓	✓
d	✓	✓	✓	✓	✓
e	✓	✓	✓	✓	✓
f	✓	✓	✓	✓	✓

CFD=0.975
(some agreement)

	h_1	h_2	h_3	h_4	h_5
a	✗	✗	✗	✓	✓
b	✗	✗	✗	✓	✓
c	✓	✗	✗	✗	✓
d	✓	✓	✗	✗	✗
e	✓	✓	✗	✗	✗
f	✗	✓	✗	✓	✗

CFD=0.333
(3 out of 5 always make the wrong decision)

classifier

input

Metrics: Coincidental Failure Diversity

	h_1	h_2	h_3		result
a	✗	✗	✗		✗
b	✗	✗	✗		✗
c	✗	✗	✗		✗
d	✓	✓	✗		✓
e	✓	✗	✓		✓
f	✓	✓	✗		✓

CFD=0.5

classifier

input

Diversity: Constructive Diversity

Diversity: Destructive Diversity

	h_1	h_2	h_3		result
a	✗	✓	✗		✗
b	✗	✓	✗		✗
c	✗	✗	✓		✗
d	✓	✗	✗		✗
e	✓	✗	✗		✗
f	✓	✗	✗		✗

CFD=0.5

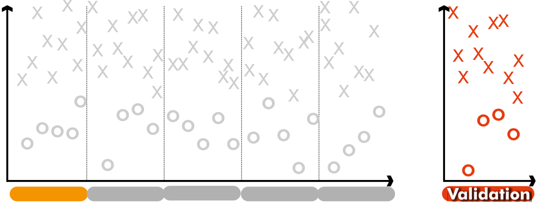
classifier

input

References

- Wang W. (2008). Some fundamental issues in ensemble methods. In *IEEE International Joint Conference on Neural Networks, 2008*, pages 2243-2250
- Kuncheva, L.I. & Whitaker, C.J. (2003). *Measures of diversity in classifier ensembles and their relationship with ensemble accuracy.* Machine Learning, 51(2):181-207

- Not exhaustive.
- Apologies to one & all for gross naivity & over-simplification.



If you intend to measure the overall accuracy/performance of your Ensemble - you must remember to set aside some data which is not used in the testing or construction of the individual members of the Ensemble.

Rooky Mistake #1