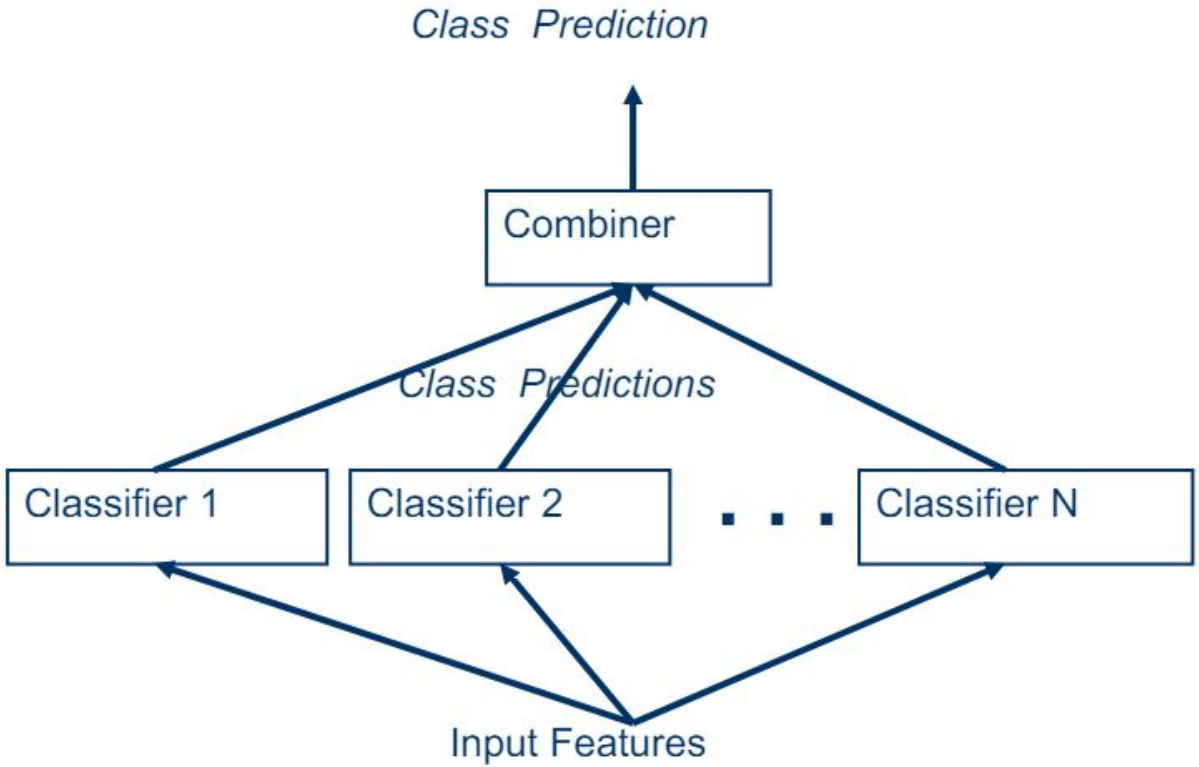
This is CS5045!

GCR:ioc7cdl

Ensemble Learning

A Classifier Ensemble



Key Questions

Which components to combine?

- different learning algorithms
- same learning algorithm trained in different ways
- same learning algorithm trained the same way

What Makes a Good Ensemble?

Krogh and Vedelsby, 1995

Can show that the accuracy of an ensemble is mathematically related:

$$\hat{E} = \overline{E} - D$$

$\hat{E}$ is the error of the entire ensemble

$\overline{E}$ is the average error of the component classifiers

D is a term measuring the diversity of the components

Effective ensembles have accurate and diverse components

Classification Fusion Techniques

- Homogenous Classifiers (Same Classifiers but different training data) e.g. Bagging, Boosting etc
- Heterogeneous Classifiers (Different Classifiers but same training data) e.g. Majority Voting, Mean etc)
- Combination of Homogenous and Heterogeneous Classifiers

Voting Techniques

- Majority Voting
- Average of Probabilities
- Product of Probabilities
- Minimum Probability
- Maximum Probability
- Median

Voting Techniques

C1	C2	C3	Majority Voting
1	1	1	1
1	1	0	1
0	0	0	0
0	0	1	0

Voting Techniques

C1	C2	C3	Average	Product	Minimum	Maximum	Median
0.9	0.5	0.5	0.63	0.23	0.5	0.9	0.5
0.5	0.5	0	0.33	0.00	0	0.5	0.5
0.1	0.1	0.1	0.10	0.00	0.1	0.1	0.1
0.4	0.4	0.6	0.47	0.10	0.4	0.6	0.4

Average = (c1+c2+c3)/3
Product = c1*c2*c3
Minimum = min(c1,c2,c3)
Maximum = max(c1,c2,c3)
Median = median(c1,c2,c3)

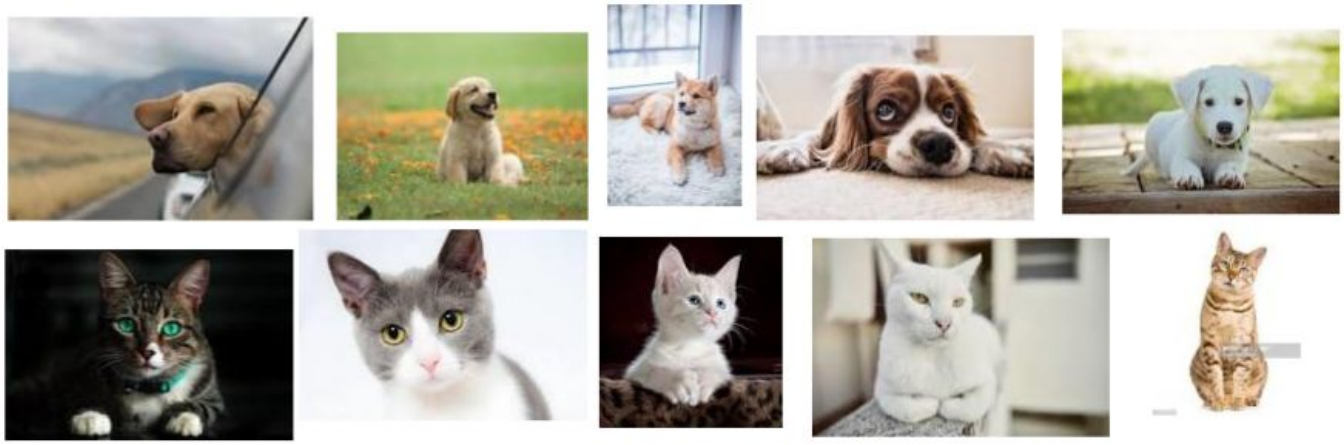
Homogenous Ensemble Classifiers

Same classifier but different training data

- Bagging
- Boosting
- Random Forest
- Others

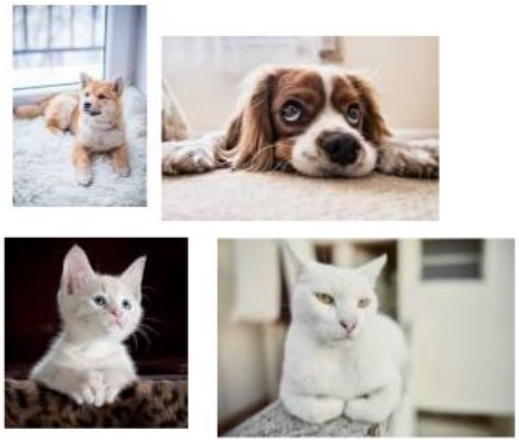
Bagging

- Sample several training sets of size *n* (instead of just having one training set of size *n*)
- Build a classifier for each training set
- Combine the classifier's predictions
- This improves performance in almost all cases if learning scheme is *unstable* (i.e. decision trees)



Full Training Data. 5 Dogs and 5 Cats

Training Dataset1: 2 Dogs and 2 Cats



Training Dataset2: 2 Dogs and 2 Cats

Boosting

- Also uses voting/averaging but models are weighted according to their performance
- Iterative procedure: new models are influenced by performance of previously built ones
  - New model is encouraged to become expert for instances classified incorrectly by earlier models
  - Intuitive justification: models should be experts that complement each other
- There are several variants of this algorithm

Stacking

- Uses *meta learner* instead of voting to combine predictions of base learners
  - Predictions of base learners (*level-0 models*) are used as input for meta learner (*level-1 model)*
- Base learners usually different learning schemes
- Hard to analyze theoretically: “black magic”

Deep Ensemble Learning

- Same concepts but applied in Deep Learning models
  - Bagging
  - Combination of Loss functions
  - Parallel models etc

Thank You

Let's Remember Each Other in Our Prayers :)

Kuch Meetha Hojaye


PICTURE
TIME

This is CS5045!
Signing Off

GCR:ioc7cdl