Carl Cortez

CIS 735 – Assignment 3

This assignment will be due by midnight on the day of live session 9

1. (40 pts) In the table below is a list of classifiers A-E, and their predicted class (1 or 2), and the weights each class would have in a weighted majority voting based on the accuracy of the classifier. ***Data fusion from asynch lectures.***

Classifier	Predicted class	Weight found empirically
A	1	0.15
B	1	0.10
C	2	0.30
D	2	0.35
E	1	0.10

What would the fused predicted class be if majority voting was used?

Majority is 1.
3 classifiers have predicted class 1; 2 classifiers have predicted class 2.

Weighted majority voting?
Majority is 2 since the combined weight for C and D is .65 and that is > .5

2. (30 pts) In class there was a discussion about the difference between decision level, score level, feature level and signal level fusion. Based on what you know from the asynchronous lectures, what type of fusion is majority voting? Weighted majority voting? Naive Bayes combination? (check videos again)

Majority voting is decision level because each classifier we see adds a tally to the number of times it has been used.

Weighted majority voting seems to be score level fusion because there is a sum being computed.

Naive Bayes could be signal level fusion.

Carl Cortez

3. (30 pts) In the live lecture, there was discussion on how to deal with missing data one involves ignoring unknowns and the other involves penalizing unknowns. Please describe the pros and cons of each method of missing data mitigation. Review the first two. Go further than what was in class and look over notes.

Pros for ignoring unknowns
- We can say some type of conclusion; the show must go on.
- We use our best judgment from what we've seen so far for inputs of other features.
- Helps acknowledge when small bits of data go missing.
- Easy to compute in the long run.

Cons for ignoring unknowns
- Can create bias if there's a saturated area that is missing.
- If not enough data is used, it could be biased.

Pros for penalizing unknowns
- Opportunity to learn from mistakes the first time.

Cons for penalizing unknowns
- More work to accommodate/penalize remaining data bits.
- Did data do something wrong or did the data collector?