Majority Voting Notes

October 14, 2010

Majority Voting

- There are three types of majority voting
  - unanimity - all classifiers agree
  - simple majority - 50% + 1 classifiers agree
  - plurality - whichever classifier has the most votes
- With Simple majority voting you need at least $\lfloor \frac{L}{2} \rfloor + 1$ classifiers to give the correct answer
  - Therefore, the accuracy of the ensemble is $P_{maj} = \displaystyle\sum_{m=\lfloor L/2 \rfloor+1}^{L} \binom{L}{m} p^m (1-p)^{L-m}$ assuming that the probability of success p is the same for all classifiers
- Classifiers, while having the same probability of success can still have markedly different distributions of error over the class labels.
- A pattern of success is a distibution of the classifier outputs such that
  - The probability of the combination of $\lfloor \frac{L}{2} \rfloor$+1 correct and $\lfloor \frac{L}{2} \rfloor$incorrect is α.
  - The probability of all L votes being incorrect is γ.
  - The probability of any other combination is zero.
- Because there are $\binom{L}{l+1}$ ways to choose $l+1$ correct out of L classifiers $p_{maj} = \binom{L}{l+1}\alpha$ which is the upper bound on the probability of an ensemble of L classifiers with probability α
- The pattern of failure alternatively is a distribution of the L classifier outputs such that
  - The probability of any combination of $\lfloor \frac{L}{2} \rfloor$ correct and $\lfloor \frac{L}{2} \rfloor + 1$ incorrect votes is β
  - The probability of all L votes being correct is δ.

- The probability of any other combination is zero.

- This results in a lower bound on the probability of the ensemble of $p_{maj} = \frac{pL-1}{l+1}$ where p is the probability of true positives for any classifier
- if each classifier has a different probability then the upper bound on the majority vote accuracy a maximum bound may be obtained by
  - ordering the classifiers by their probability such that $p_1 \leq p_2 \leq ... \leq p_l$
  - Let $k = \frac{L+1}{2}$
- Then the upper bound is $\max P_{maj} = \min\{1, \Sigma(k), \Sigma(k-1), ..., \Sigma(1)\}$ where $\Sigma(m) = \dfrac{1}{m} \displaystyle\sum_{i=1}^{L-k+m} p_i$.
- The resulting lower bound is similarly $\min P_{maj} = \max\{0, \xi(k), \xi(k-1), ..., \xi(1)\}$where $\xi(m) = \frac{1}{m} \displaystyle\sum_{i=k-m+1}^{L} p_i - \dfrac{L-k}{m}$.

Weighted Majority Voting

- Not all classifiers are created equal
  - Use weighted majority votes when classifiers do not all have the same accuracy rates
  - This gives better classifiers more power in the decision-making process
- Label outputs like so

$$d_{i,j} = \begin{cases} 1 & \text{if } D_i \text{ labels x in } w_j \\ 0 & \text{otherwise} \end{cases}$$

- Obtain discriminant function for class w_j
  - Let b_i be the coefficient for classifier D_i
  - $g_j(x) = \displaystyle\sum_{i=1}^{L} b_i d_{i,j}$
  - In English: the value of the discriminant function is the sum of the coefficients for those members whose output at x is w_j

- How do we normalize coefficients?
  - As conveniently as possible
  - Normalize coefficients to 1
  - $\sum_{i=1}^{c} b_j = 1$
- Maximizing accuracy of P_{maj}^{w} by assigning weights like so
  - $b_i \propto \log \frac{p_i}{1-p_i}$
- The discriminant function can be reduced to: $g_i(x) = \log P(w_j) + \sum_{i=1}^{c} d_{i,j} \log \frac{p_i}{1-p_i}$