Supplementary materials of article “An Analysis of Diversity Measure” submitted to *Machine Learning Journal*, 2006.

This supplementary file contains two sections. The first one presents bigger versions of Figs 2.1-2.6. The second part presents the derivations of two additional diversity measures that are also relevant to the work presented in the paper.

Section 1

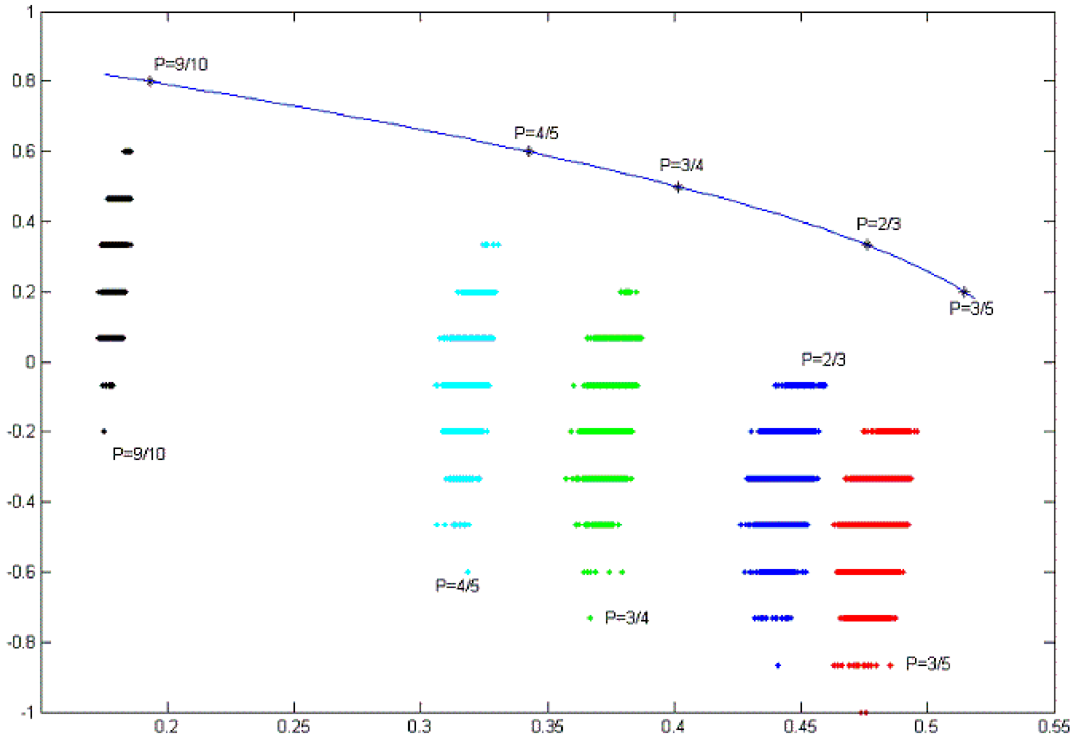

Figure 2.1: Disagreement measure

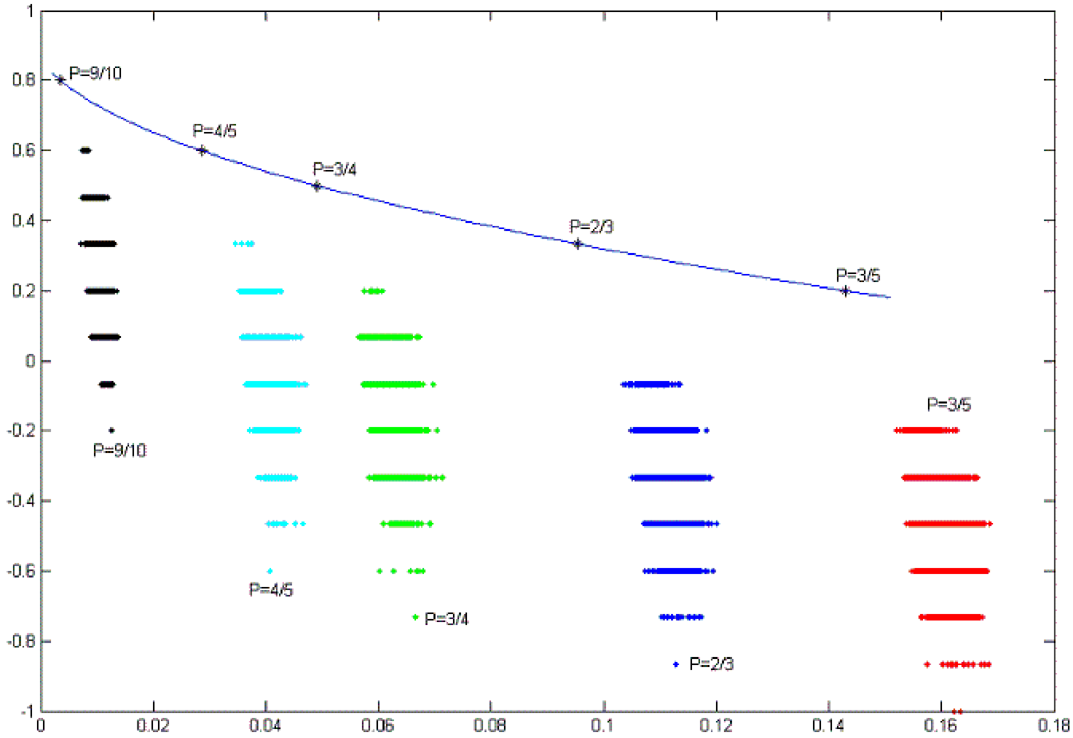

Figure 2.2: Double-fault measure

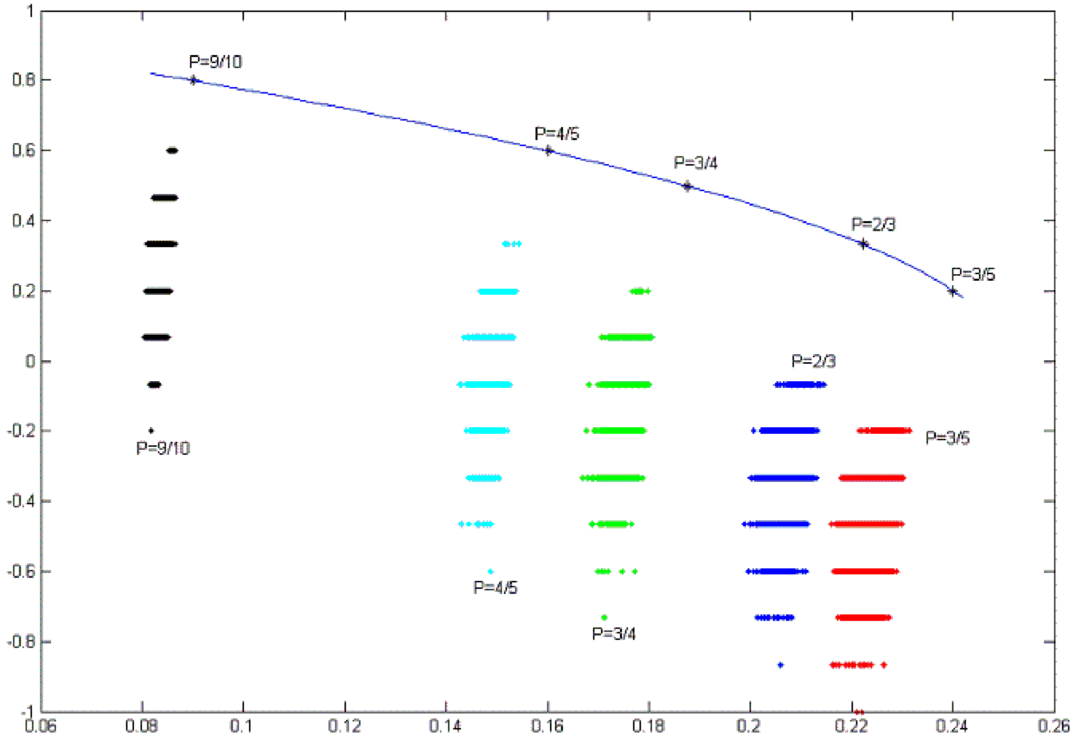

Figure 2.3: *L*=15, Kohavi-Wolpert variance

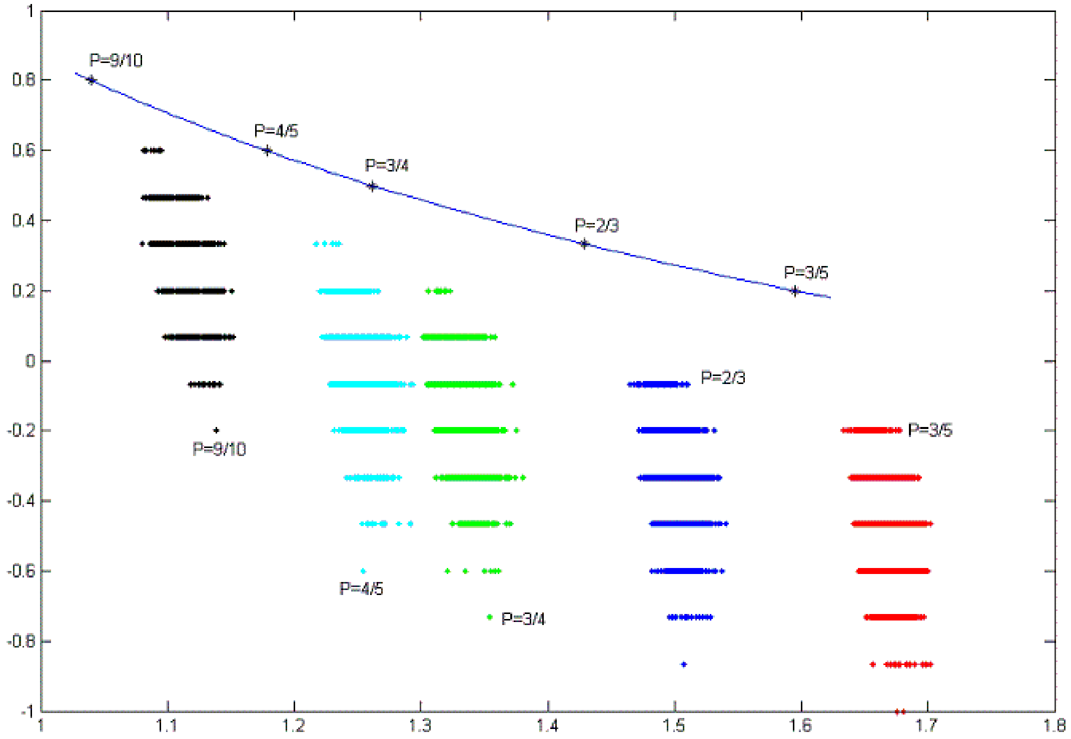

Figure 2.4: Measurement of interrater agreement

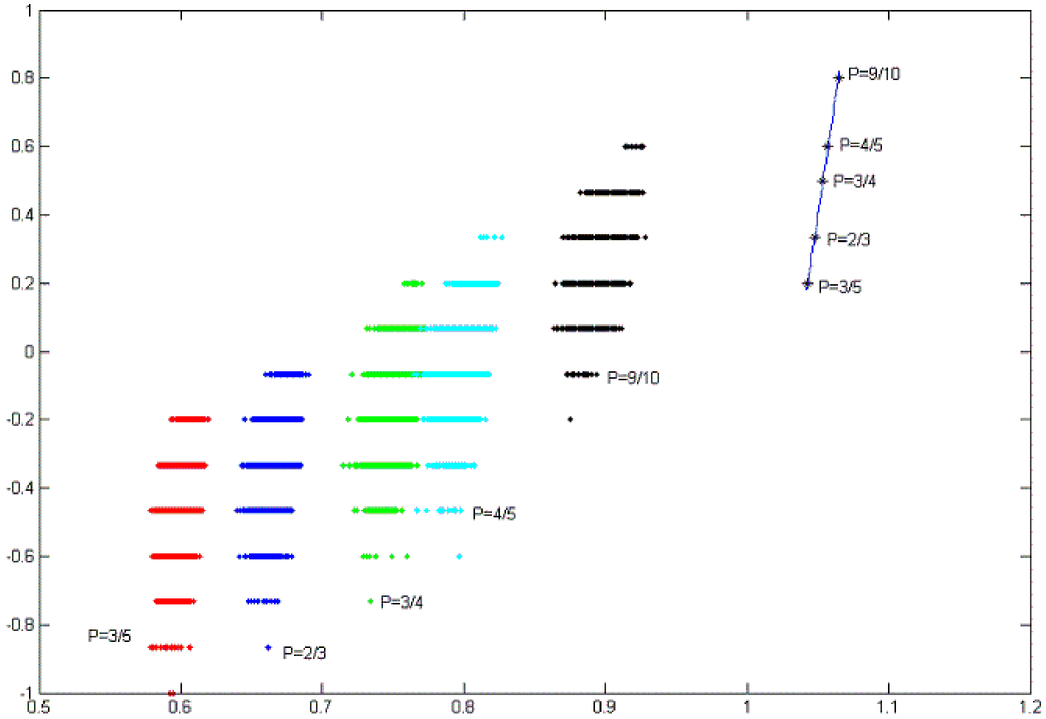

Figure 2.5: Generalized diversity

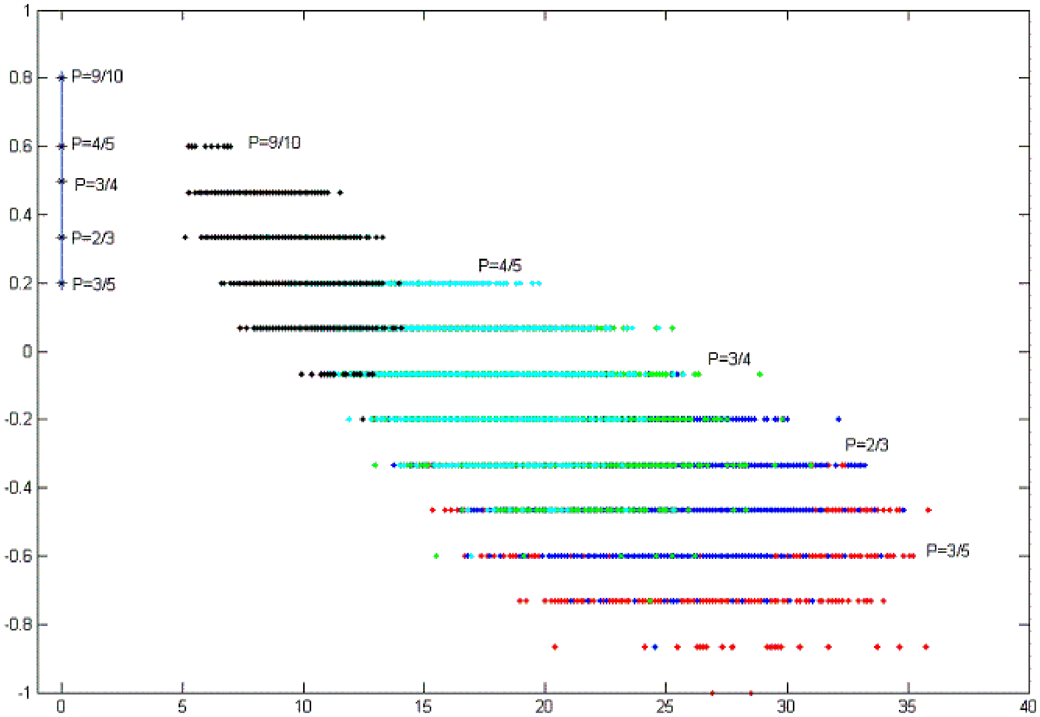

Figure 2.6:Measure of "Difficulty"

Fig.2: The figures show the relationships between diversity measures and the minimum margin. *L*=15, horizontal axis represents the diversity measures and vertical axis represents the minimum margin. The upper bound of the minimum margin is also shown in the figures corresponding to *P*=3/5, 2/3, 3/4, 4/5, and 9/10 with "*".

Section 2

Two more relevant diversity measures

The entropy measure *E*

Kuncheva et al. (2003a) claimed that the highest diversity among base classifiers could be manifested by the equation:

$$E = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\left(L-\lceil L/2 \rceil\right)} min\{l_i, L-l_i\}. \tag{1}$$

The diversity increases with increasing values of the entropy, and the minimum value of this measure is 0.

Since: $min\{l_i, L-l_i\} + max\{l_i, L-l_i\} = L$

and $max\{l_i, L-l_i\} - min\{l_i, L-l_i\} = \left|L-2l_i\right|$,

Equation (1) can be re-written as:

$$E = \frac{1}{N\left(L-\lceil L/2\rceil\right)}\sum_{i=1}^{N}\frac{L-\left|L-2l_i\right|}{2}$$

$$= \frac{L}{2\left(L-\lceil L/2\rceil\right)} - \frac{1}{2N\left(L-\lceil L/2\rceil\right)}\sum_{i=1}^{N}\left|L-2l_i\right| \qquad (2)$$

Since $\sum_{i=1}^{N}\left|L-2l_i\right| \geq \left|\sum_{i=1}^{N}\left(L-2l_i\right)\right|$, we can get:

$$E \leq \frac{L}{2\left(L-\lceil L/2\rceil\right)} - \frac{1}{2N\left(L-\lceil L/2\rceil\right)}\left|\sum_{i=1}^{N}\left(L-2l_i\right)\right|$$

$$\leq \frac{L}{2\left(L-\lceil L/2\rceil\right)} - \frac{1}{2N\left(L-\lceil L/2\rceil\right)}\left|NL-2NL\left(1-P\right)\right|$$

$$\leq \frac{L}{2\left(L-\lceil L/2\rceil\right)} - \frac{L}{2\left(L-\lceil L/2\rceil\right)}\left|2P-1\right|. \qquad (3)$$

$$max\left(E\right) = \frac{L}{2\left(L-\lceil L/2\rceil\right)} - \frac{L}{2\left(L-\lceil L/2\rceil\right)}\left|2P-1\right| \qquad (4)$$

when $L-2l_i \geq 0 \ \ \forall i$. (5)

Equation (5) shows that maximizing *E* does not require the uniformity condition to be satisfied. However, it requires l_i to be smaller than *L*/2 for any *i*, which represents a relatively balanced distribution of l_i (or in other words, requires $min(m_i) \geq 0 \ \forall i$). Hence, there still exists some implicit connection between the entropy measure and the uniformity condition. Further, we observe two more properties of the entropy measure from equations (4) and (5). First, given *P*, equation (5) can be satisfied by many different distributions of l_i. Different distributions of l_i usually correspond to different generalization performance, but the entropy measure fails to differentiate them. Second, since *E* is also parameterized by *P*, the maximum value of *E* can only be achieved with *P*=0.5. These two properties are neither desirable nor reasonable, and thus the entropy measure may not be suitable.

Coincident failure diversity

Proposed by Partidge and Krzanowski (1997), coincident failure diversity is a modification of the generalized diversity. For each sample in the training set, let T_j denote the probability that l_i=j. The coincident failure diversity is defined as:

$$CFD = 0 \text{ if } T_0 = 1,$$

$$CFD = \frac{1}{1-T_0}\sum_{j=1}^{L}\frac{L-j}{L-1}T_j \text{ if } T_0 < 1. \qquad (6)$$

Let $n(j)$ be the number of samples that are classified incorrectly by j base classifiers. Since $T_0 + \sum_{j=1}^{L} T_j = 1$ and $\sum_{j=1}^{L} jn(j) = \sum_{i=1}^{N} l_i$, then when $T_0 < 1$,

$$\begin{aligned}
CFD &= \frac{1}{1-T_0}\left(\frac{L}{L-1}\sum_{j=1}^{L}T_j - \frac{1}{L-1}\sum_{j=1}^{L} jT_j\right)\\
&= \frac{1}{L-1}\left(L - \frac{1}{1-T_0}\sum_{j=0}^{L} jT_j\right)\\
&= \frac{1}{L-1}\left(L - \frac{N}{N-n(0)}\sum_{j=0}^{L}\frac{jn(j)}{N}\right)\\
&= \frac{1}{L-1}\left(L - \frac{1}{N-n(0)}\sum_{j=0}^{L} jn(j)\right)\\
&= \frac{1}{L-1}\left(L - \frac{1}{N-n(0)}\sum_{j=0}^{L} l_i\right)\\
&= \frac{1}{L-1}\left(L - \frac{NL(1-P)}{N-n(0)}\right)\\
&= \frac{L}{L-1}\cdot\frac{NP-n(0)}{N-n(0)} \qquad (7)
\end{aligned}$$

If we regard P as a constant, the uniformity condition is not required by equation (7). But similar to the entropy measure, equation (7) is maximized when n(0)=0, which also implicitly represents a balanced distribution of l_i (or in other words, requires $\max(m_i) < 1\ \forall i$). Since P is a variable, equation (7) is maximized when l_i=1 $\forall i$, which actually is a special case of the uniformity condition.

References:

Kuncheva, L., & Whitaker, C. (2003a). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, *51*, 181-207.

Patridge, D., & Krzanowski, W. J. (1997). Software diversity: Practical statistics for its measurement and exploitation. *Information & Software Technology*, *39*, 707-717.