Problem 1: No Free Lunch Theorem

1. For a particular algorithm h and a given training set D, the expected error over all two-category problems can be represented as:

$$\epsilon[E|D,h] = \sum_F \sum_{x \notin D} P(x)[1 - \delta(F(x), h(x))]P(F|D)$$

where we suppose there are two categories ω_1, ω_2, if the majority in the dataset D is ω_1, we will have $h(x) = \omega_1$ for all x, thus, $1 - \delta(F(x), h(x)) = 1$ iff $F(x) = \omega_2$. Therefore:

$$\epsilon[E|D,h] = \sum_F \sum_{x \notin D} P(x)P(F(x) = \omega_2|D)$$

Since we have no prior knowledge concerning $P(F|D)$, $\sum_F P(F(x) = \omega_2|D) = 0.5$, thus we will have

$$\epsilon[E|D,h] = \sum_F \sum_{x \notin D} P(x)P(F(x) = \omega_2) = \sum_{x \notin D} P(x) \sum_F P(F(x) = \omega_2|D) = 0.5$$

Similarly, if the majority in D is ω_2, we can still prove that $\epsilon[E|D,h] = 0.5$.

Therefore, the averaged over all two-category problems of a given number of features, the off-training set error is 0.5.

2. Similar to the previous step, if the majority in the dataset D is ω_2, we will have $h(x) = \omega_1$ for all x, thus, $1 - \delta(F(x), h(x)) = 1$ iff $F(x) = \omega_2$. Therefore:

$$\epsilon[E|D,h] = \sum_F \sum_{x \notin D} P(x)P(F(x) = \omega_2|D)$$

Since we have no prior knowledge of $P(F|D)$, $\sum_F P(F(x) = \omega_2|D) = 0.5$, thus we will have

$$\epsilon[E|D,h] = \sum_F \sum_{x \notin D} P(x)P(F(x) = \omega_2) = \sum_{x \notin D} P(x) \sum_F P(F(x) = \omega_2|D) = 0.5$$

Similarly, if the majority in D is ω_1, we can still prove that $\epsilon[E|D,h] = 0.5$.

Therefore, the averaged over all two-category problems of a given number of features, the off-training set error is 0.5.

3. For any fixed training set D, both the majority learning algorithm and the minority learning algorithm have the same average off-training set error as 0.5, which is the same as random guess. For any given learning algorithm, the result can be regarded as a combination of these two algorithms: for some test data, the output is the same as the majority algorithm, while for others, the output is the same as the minority algorithm. If F is uniformly distributed, this illustrated the part 2 of the No Free Lunch Theorem:
For any fixed training set D, uniformly averaged over F, $\epsilon_1(E|F,D) - \epsilon_2(E|F,D) = 0$.