ABSTRACT. In spite of the general wisdom that much more examples need to test than to train a neural network, vice-versa we show that testing the approximation capability of a neural network generally demands smaller sample size than training it.

We move in an extended PAC learning framework and use some recent results in terms of sentry functions of a concept class to statistically proof our claims.

1. Introduction
PAC learning theory [6, 9] gives exact conditions for having that, given:
- a class G of functions g: R->R
- a loss function l(y,g(x))
- a distribution law P on R and a sample Σ=((X_1, g(X_1)),..., (X_m, g(X_m)))
- a pair of real numbers ε and δ close to zero,
a function h is computable such that

\[ p(m)(E[l(h(X,L),g(X))] \leq \varepsilon) \geq 1 - \delta, \]

where \( p(m) \) is the probability measure in the product space \( R^m \).

One of these conditions concerns the sample complexity of the inference problem, and generally consists of upper and lower bounds on the size \( m \) of \( \Sigma \). These bounds, that come from a clever and sophisticate theory [10], are however only seldom used by those people which want to learn \( h \) through a neural network. The actual size of the training set is of one or more order less than the size of \( \Sigma \) requested by the sample complexity, where this discrepancy can be attributed to the following inner limitations of the above theory:

1. the bounds on \( m \) constitute worst case results
2. the theory is not yet tailored to take into account the peculiarities of the neural network and of the learning algorithm.

Thus, we generally disregard the bounds, use a short training set to train the neural network and try to check the adequacy of this set by testing the inferred \( h \) on a new set of example (the test set). If the network (i.e. \( h \)) performs on the test nearby as well as on the training set we are satisfied and declare the learning task successful.

To be safe, we are used to employ very large test sets mainly because their processing is relatively inexpensive and, of course, more is better than less.

What news can testing gives us that we did not yet know from training? We answer this question in sect. 2 as it concerns boolean functions, after the question is framed in the sentry function theory [2]. In sect. 3 we extend our results to real functions. Our conclusion is that it is generally wasteful to use a test set larger than a training set, where a suitable ratio between the two sizes can be computed if some complexity parameters of the learning task are available.

2. Degrees of freedom in learning and checking a boolean function.
Let us start with the distribution free case and come us to consistent statistics.
**Definition 1:** Given a set $X$, a *class of formulas* or *class of concepts* $C$ is a set of boolean functions $c$ on $X$. These boolean functions are also called concepts $c$. By abuse of notation, we shall not make any distinction between $c$ and its support, i.e. the set of elements $x$ such that $c(x) = 1$. Therefore $C$ can be also view as a set of subsets of $X$.

Willing infer an approximation $h$ for $c$, let call it hypothesis, we distinguish inside the sample $\Sigma$ a set of points which constitute the pivots around which to draw that hypothesis or other hypotheses equivalent in a proper metric space. Actually, they act as sentinels that forbid the expansion of $c+h$ (symmetric difference between $c$ and $h$). Thus, in figure 1 $x_1$ and $x_2$ outside $c$ are sufficient to prevent that a larger circle non containing them includes $c$.

The remaining sampled points constitute then rear-guard which, if numerous and fairly scattered on $X$, give confidence that each pivot has been considered. Linking pivots, confidence and sample size allows us to bind the learning sample complexity.

![Figure 1](image-url)

*Figure 1* The two points $x_1$, $x_2$ are sufficient to sentinel $c$

**Definition 2:** Given a concept class $C$ on $X$, an *outer sentry* function on $C$ is a total function $S : C \cup \{\emptyset, X\} \to 2^X$ satisfying the following conditions:

1. the elements of $S(c)$ are outside $c$, i.e.: $c \cap S(c) = \emptyset$.
2. if $c_1, c_2 \in C$, $c_2 \not\subseteq c_1$ and $c_1 \cup S(c_1) \subseteq c_2 \cup S(c_2)$ then $c_2 \cap S(c_1) \neq \emptyset$.
3. no $S' \neq S$ exists satisfying (1) and (2) and having the property that $S'(c) \subseteq S(c)$ for every $c$.
4. whenever $c_1$ and $c_2$ are such that $c_1 \subseteq c_2 \cup S(c_2)$ and $c_2 \cap S(c_1) = \emptyset$, then the restriction of $S$ to $C - \{c_2\}$ is a sentry function on $C - \{c_2\}$.

Terminology: $c^+ = c \cup S(c)$. $S(c)$ is the *outer frontier* upon $S$ of $c$, their elements are called sentry points. $c_2$ is *sentinelled* by $S(c_1)$ iff $c_2 \cap S(c_1) \neq \emptyset$.

**Remark 1:** A frontier does not fully identify the sentinelled concept. A given concept class might admit more than one outer sentry function. Condition (4) prevent us from building sentry functions which are *unnatural*, where some sentry points of $c_1$ have the sole role of artificially increasing the elements of $c_1^+$ in order