Machine learning 2 - Final Assignment

Frans de Boer
frans.deboer
5661439

27 juni 2022

1 PAC Learning

1.a Extend the proof that was given in the slides for the PAC-learnability of hyper-rectangles: show that axis-aligned hyper-rectangles in n-dimensional feature spaces (n > 2) are PAC learnable.

We can extend the proof by taking the number of sides/faces of the hyper-rectangle into account.

- A hyper-rectangle in 2-dimensional space is simply a rectangle and has 4 sides.
- A hyper-rectangle in 3-dimensional space is a rectangular box and has 6 sides.

For now I will assume the number of sides on a n-dimensional hyper-rectangle to be $F(n)$, and I will get back later to the calculation of $F(n)$.

To start we can follow the steps from the lecture. We have the ground truth n-dimensional hyper-rectangle R, and the current tightest fit rectangle R'. The error $(R-R')$ can be split into $F(n)$ strips T'. On each of these strips we can now grow a new strip T such that the probability mass of T is $\frac{\varepsilon}{F(n)}$.

If T covers T' for all strips then the probability of an error is

$$P[error] \leq \sum_{i=0}^{F(n)-1} P[T_i] = F(n)\frac{\varepsilon}{F(n)} = \varepsilon \tag{1}$$

We can now estimate the probability that T does not cover T'

$$P[random\ x\ hits\ T] = \frac{\varepsilon}{F(n)}$$
$$P[random\ x\ misses\ T] = 1 - \frac{\varepsilon}{F(n)}$$
$$P[m\ random\ x\text{'s}\ misses\ T] = (1 - \frac{\varepsilon}{F(n)})^m$$

Since we have $F(n)$ strips:

$$P[\textit{m random x's miss any Ts}] \leq F(n)(1 - \frac{\varepsilon}{F(n)})^m$$
$$P[\textit{R' has larger error than } \varepsilon] \leq F(n)(1 - \frac{\varepsilon}{F(n)})^m \ < \delta$$

Bounding the chance that our R' has an error larger than ε by δ:

$$F(n)(1 - \frac{\varepsilon}{F(n)})^m \ < \delta$$

Using $e^{-x} \geq (1-x)$

$$F(n)e^{-m\varepsilon/F(n)} \geq F(n)(1 - \frac{\varepsilon}{F(n)})^m$$

So instead we can use:

$$F(n)e^{-m\varepsilon/F(n)} < \delta$$
$$-m\varepsilon/F(n) < \log(\delta/F(n))$$
$$m\varepsilon/F(n) > \log(F(n)/\delta)$$
$$m > (F(n)/\varepsilon)\log(F(n)/\delta)$$

So any n-dimensional hyper-rectangle is learnable.

1.b Assume we have a 2-dimensional feature space R2, and consider the set of concepts that are L1-balls: $c = \{(x,y) : |x| + |y| \leq r\}$ **(basically, all L1-balls centered around the origin). Use a learner that fits the tightest ball. Show that this class is PAC-learnable from training data of size m.**

Same thing? what. literally just question a but now rotated 45 degrees

1.c Now we extend the previous class by allowing the varying center: $c = \{(x,y) : |x - x0| + |y - y0| \leq r\}$**. Is this class PAC-learnable? Proof if it is, or otherwise disproof it.**

maybe, maybe not. can't you still just do the tightest fit L1 ball? consistent learner so PAC learnable?

2 VC Dimension

Let us assume we are dealing with a two-class classification problem in d-dimensions. Let us start off with refreshing our memories. Consider the class of linear hypotheses (i.e., all possible linear classifiers) in d-dimensional space.

2.a Given N data points, how many possible labelings does this data set have?

Each data point can have 2 labels, so 2^n

2.b What is the dimensionality d we need, at the least, for our class of linear hypotheses to be able to find perfect solutions to all possible labelings of a data set of size N? Stated differently, what is the smallest d that allows us to shatter N points?

d - 1?

2.c Consider d = 1 and a data set of size N. At maximum, how many different labelings of such a data set can decision stumps solve perfectly, i.e., with zero training error?

This would be all cases where there is no overlap.