

Problem Set 4

Chen Zhang

Handed In: 10/17/2014

1 Problem 1

The VC dimension of a circle is 3. When there are three points(not in a line), the most difficult case is when the points with the largest distance, denoted A and B, have the same label, and remaining point, denoted C, has another label. In this case, connect the two points with the largest distance, call the line ab. Pick any point between line ab and point C, denote it as D. In this case draw a circle with A B D point(three points form a circle), it is obvious that this circle has A and B in it but does not have C in it. So three points can be shattered by a circle.

If there are four points, when any of the three points are on the same line, the circle cannot shatter it. When no three points are on the same line, the four points form a quadrilateral. In this case, the center of the circles that contains the diagonal points is on the middle Vertical of the diagonal line. In this case, at least one of the two middle Vertical has the following property : any point on this line has smaller distance to the other two points than to the diagonal points of which this line is a middle Vertical. This means that at least one circle which contains the two points, also contains the other two points. This means that a circle cannot shatter four points.

Therefore, the VC dimension of a circle is three.

1.1 Part 2

The VC dimension of k interval is $2k$. It is easy to see that for any $2k$ points, it can be shattered by k interval with the proposed condition. However, if there are $2k+1$ points, each labeled $(-1)^{i+1}$, then there would be $k+1$ number of separate positive points. However, we only have k interval, meaning we can only determine k separate positive points. Thus this $2k+1$ points can not be shattered by k interval with the proposed condition. Thus the VC dimension of k interval is $2k$.

2 Problem 2

2.1 Part 1

If a concept c can be expressed as $c = \langle (c_1, b_1), \dots, (c_\ell, b_\ell), b \rangle$, then its negated concept $\neg c$ can be expressed as $\neg c = \langle (c_1, \neg b_1), \dots, (c_\ell, \neg b_\ell), \neg b \rangle$. The proof is simple: the disjunction that this data satisfies has the negated label. If this data use the default value, this default value is also negated.

2.2 Part 2

By construction, it is easy to observe that a K-DNF can be readily transformed to a K-DL. Assume a K-DNF concept d has concept as $d = \langle (d_1, b_1), \dots, (d_\ell, b_\ell) \rangle$. Then it could be transformed into a k-DL c as $c = \langle (d_1, 1), (d_2, 1), \dots, (d_\ell, 1), 0 \rangle$. It means that to put a K-DNF in the form of a K-DL, we make all the labels in the DL 1 and the default label 0. Since this condition is strict, meaning that some k-DL can not be represented as a K-DNF. Thus $k\text{-CNF} \subseteq k\text{-DL}$.

It is widely known that a k-CNF can be transformed into a corresponding k-DNF. Thus $k\text{-CNF} \subseteq k\text{-DL}$.

Therefore we have $k\text{-CNF} \subseteq k\text{-DL}$, and $k\text{-CNF} \subseteq k\text{-DL}$. Therefore $k\text{-DNF} \cup k\text{-CNF} \subseteq k\text{-DL}$.

2.3 Part 3

The algorithm follows:

Now we have all the data, m being the number of data in the data set. Each data has the form $X_1, X_2, \dots, X_n, \langle 0, 1 \rangle^n$. For the first attribute, choose the first literal of all the existing data and see if this condition holds: only the data with one same label would satisfy this literal. If no literal satisfy this condition, don't do anything. If there is literal that satisfies this condition, take this literal as the first disjunction in the decision list, its label b_j the corresponding label and cross out the data that satisfies this disjunction from the data set S . Total number of cases considered is smaller than $m \times C_1^1$.

Then move to the first two attributes. Choose combinations of the first two existing literals in the data set as a conjunction such that only data with one same label would satisfy one of these combinations. If no combination meets the condition that only data with one same label would satisfy it, don't do anything. If there are combinations of literals such that only data with the same label would satisfy, take this combination as the second conjunction (and the third) in the decision list and their label as the corresponding label in the decision list. Then cross out the data that satisfies this conjunction from the data set S . Number of cases considered is smaller than $m \times (C_2^1 + C_2^2)$.

Continue this process until all n attributes are considered. In this step the number of cases considered is smaller than $m \times (C_n^1 + C_n^2 + \dots + C_n^n)$. The labels of the last data (maybe more than one data has this same label) can be taken as the default value. This process is guaranteed to find a decision list that is consistent with the data set because by nature the data that satisfy the later conjunction do not satisfy the previous conjunction. Besides, the worst case is that this decision list is simply a record of all the data and its corresponding label. Thus it is consistent with the existing data set.

2.4 Part 4

We know that the total cases N considered using the above procedure is smaller than $N = m \times [(C_2^1 + C_2^2) + (C_3^1 + C_3^2 + C_3^3) + (C_4^1 + C_4^2 + C_4^3 + C_4^4) + \dots + (C_n^1 + C_n^2 + \dots + C_n^n)]$. We know that C_n^k can be taken as polynomial and that a finite number of combination of polynomials is still

a polynomial. Thus the total cases considered is polynomial with the number of attributes n , and runs in polynomial time in m for any $k \geq 1$.

We have that any decision list with n variables can be encoded in $(kn^k \log n)$, and by Occam's Razor the sample size bound is $m = O(\frac{1}{\epsilon}(\log \frac{1}{\delta} + kn^k \log n))$

Also since the described procedure runs in time polynomial in m (calculated above), we get efficient PAC learning for k -DL.

3 Problem 3

3.1 Part 1

Imagine that each term is exactly of k literals. In this case, since we are dealing with DNF, in the entire concept space, only the combination of k same literals in x_1 and x_2 contributes to the value of the kernel. In this case, the kernel value for this x_1 and x_2 is $C_{same(x_1, x_2)}^k$. This function is linear in n since picking the same value in x_1 and x_2 grows linear in the dimension of x . Now if we expand to the space of $1 - k$, we have the kernel value as $K(x_1, x_2) = C_{same(x_1, x_2)}^1 + C_{same(x_1, x_2)}^2 + C_{same(x_1, x_2)}^3 \dots C_{same(x_1, x_2)}^k$. Calculating $C_{same(x_1, x_2)}^k$ is a linear of n and $K(x_1, x_2)$ is a finite fixed number of combination of the time to calculate $C_{same(x_1, x_2)}^k$. Therefore $K(x_1, x_2)$ is a linear of n .

3.2 Part 2

The algorithm follows:

Initilize summation=0. The data label to be determined is x_i .

Loop over all the data, call it x_j

Call $S(x_j) = Label(x_j)$, calculate $K(x_i, x_j)$ based on the formula proposed in part 1.

Then summation=summation+ $S(x_j) + K(x_i, x_j)$.

After the loop is over, assign the label to x_i based on the sign of the summation.

3.3 Part 3

Remember that for normal Perceptron, the bound is $\frac{R^2}{\gamma^2}$, in which R is the bound of norm of x_i and γ can be considered as the margin of the data set. In our case, we have a expanded space. Doing the calculation in the expanded space, call the mistake bound as $\frac{R'^2}{\gamma'^2}$. R' is the bound of norm of x_i in the expanded space, denoted as t_i . γ can be considered as the margin of the data in the t space.

Making use of the property of K-DNF, we know that the data which gives negation is $(0, 0, 0, \dots, 0)$ in the expanded space. And the data which gives positive value but closest to the negated value is $(0, 0, 1, 0, \dots, 0)$ (one 1 and the other attributes are 0), no matter what concept it uses. Thus we know that the margin $\gamma' = \frac{1}{2}$ with $\theta = 0$. If $\theta \neq 0$, then the margin is $\frac{1}{2} - \theta$. Thus the mistake bound is $\frac{R'^2}{\gamma'^2}$, in which R' is the bound of norm of t_i (x_i in the expanded space t) and $\gamma' = \frac{1}{2} - \theta$