Problem Set 4 for Machine Learning 15 Fall

Jingyuan Liu
AndrewId: jingyual
`jingyual@andrew.cmu.edu`

1 VC dimension

1.1 Show the VC dimension of linear classifier

To prove that the linear classifier h with x in R^n has the VC dimension of n + 1, we need to prove that $VCdim(h_n) >= n + 1$, and then prove that $VCdim(h_n) <= n + 1$.

(a). Prove $VCdim(h_n) >= n + 1$

First of all, we could use the Mathematical Induction to prove that $VCdim(H) >= n + 1$:

For n = 1, it is easy to get $VCdim(h_1) >= 2$

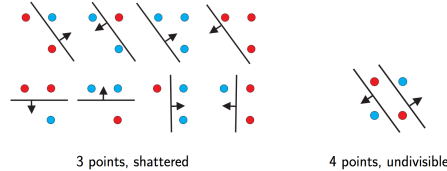For n = 2, we could also get $VCdim(h_2) >= 3$, which could be proved using following figures:

3 points, shattered   4 points, undivisible

Figure 1: The representation of $VCdim(h_2)$

For n = i, we assume that the $VCdim(h_i) >= i + 1$. We could form a matrix of (i+1)*(i+1), with the rank of i+1, since the VCdim is i+1.

For n = i+1, we could form a matrix of (i+2)*(i+2), since the variable x have one more independent dimension. Therefore, $VCdim(h_{i+1}) >= VCdim(h_i) + 1 >= i + 2$.

Therefore, we have $VCdim(h_n) >= n + 1$.

(b). Prove $VCdim(h_n) <= n + 1$

Then we should prove that $VCdim(h_n) <= n + 1$:

Suppose we could shatter n+2 points, then using the convex combinations of points in C, we could seperate points into S1 and S2, that:

$$conv(s1) \cap conv(s2) \neq \emptyset \tag{1}$$

However, if $VCdim(h_n) = n + 2$, then we could spilt points to half space contains s1 and the complement of the half space contains s2. This implies that the both half space contains the convex hull of s1 and the complement of the half space the contains the convex hull of s2. Thus, we get:

$$conv(s1) \cap conv(s2) = \emptyset \tag{2}$$

These two observations are contradictory to each other. So we know that for the n+2 points here, we could not use the linear classifier to shatter all of them.

In conclusion, we prove that $VCdim(h_n) >= n + 1$ and $VCdim(h_n) <= n + 1$ for the linear classifier, then we have $VCdim(h_n) = n + 1$

1.2 Show the VC dimension of axis-aligned boxes

Similarily, for this case, to prove that the axis-aligned boxes classifier h with x in R^n has the VC dimension of 2n, we need to prove that $VCdim(h_n) >= 2n$, and then prove that $VCdim(h_n) <= 2n$.

(a). Prove $VCdim(h_n) >= 2n$

First of all, we need to prove that for any case, $VCdim(h_n) >= 2n$, which means if the x has n dimensions, we can use the axis-aligned boxes classifiers to shatter 2n points.

Suppose we have n dimensions, then we can map all the x to a dimension i. In the dimension i, we can have x^i_{max} and x^i_{min}. Therefore, in this dimension i, we could always shatter at least 2 points via a split between the x^i_{max} and x^i_{min}.

Considering we have n dimensions, and the mapping of each dimension of the data is independent to the mapping of other dimension of the data. Therefore, we could at least shatter 2*n points via the axias-alighed boxes. Therefore, we have $VCdim(h_n) >= 2n$.

(b). Prove $VCdim(h_n) <= 2n$

Then we need to prove that $VCdim(h_n) <= 2n$. Suppose we have 2n + 1 points, we could always find the 2n "boundries", which is the min value and max value for each dimension, and form an "area". Then the 2n+1th points will be guaranteed to appear within the selected "area".

For any 2n+1 points, we could always transfer to the above mentioned scenario. In this scenario, we could not classify the 2n+1th point, because it is contained within the "area", and no rules are guaranteed to rightly classify it.

Therefore, by mapping the data to each dimension and form an "area", we could prove that $VCdim(h_n) <= 2n$.

In conclusion, we prove that $VCdim(h_n) >= 2n$ and $VCdim(h_n) <= 2n$ for the axis-alighed boxes, then we have $VCdim(h_n) = 2n$

2 AdaBoost

3 Gaussian Mixture Model

3.1 Show the expectation

$$E(x) = \int p(x)dx = \sum_k p(x_k)x_k \tag{3}$$

$$= \sum_k \pi_k E_k(x) \tag{4}$$

We know that for each k, the implict distribution of x is a Gaussian distribution, and $E_k(x) = \mu_k$, μ_k is the mean of the kth guassian distribution, so we have:

$$E(x) = \sum_k \pi_k \mu_k \tag{5}$$

3.2 Show the covariance

$$cov(x) = E(xx^T) - E(x)E(x)^T \tag{6}$$

$$= \sum_k \pi_k E_k(xx^T) - E(x)E(x)^T \tag{7}$$

$$= \sum_k \pi_k (\Sigma_k + \mu_k \mu_k^T) - E(X)E(x)^T \tag{8}$$

4 K-Means

5 Collaboration

I dicussed with Zheng Chen with problem 1.2 on how to calculate the parameters, and he also informed me about 2.2.1(b).