Comp 6321 - Machine Learning - Assignment 4

Federico O'Reilly Regueiro

December 4^{th}, 2016

Question 1: VC dimensions

1.a $[a, \infty)$

We can shatter a single point p_0, $p_0 \in \mathbb{R}$.:

point	label	h
p_0	$\oplus$	$[a, \infty), a < p_0$
p_0	$\ominus$	$[a, \infty), a > p_0$

But if we have two points, $p_0, p_1 \mid p_0 < p_1, p_0 \in \oplus, p_1 \in \ominus$, then $[a, \infty)$ cannot shatter them. Therefore, for this class of hypothesis: $VC_{dim} = 1$

1.b $(-\infty, a]$ or $[a, \infty)$

Similarly to the previous question, we can shatter one point. Additionally, we can shatter two points, $p_0, p_1 \mid p_0 < p_1, p_0$:

point	label	h
p_0	$\ominus$	$(-\infty, a], a < p_0$
p_1	$\ominus$	
p_0	$\ominus$	$[a, \infty), p_0 < a < p_1$
p_1	$\oplus$	
p_0	$\oplus$	$(-\infty, a], p_0 < a < p_1$
p_1	$\ominus$	
p_0	$\oplus$	$[a, \infty), a < p_0$
p_1	$\oplus$	

However, three points $p_0, p_1, p_2, \mid p_0 < p_1 < p_2, p_0 \in \ominus, p_1 \in \oplus, p_2 \in \ominus$ cannot be shattered. Therefore, for this class of hypothesis: $VC_{dim} = 2$

1.c Finite unions of one-sided intervals

The union of more than one left-side interval $(-\infty, a] \cup (-\infty, b] \ldots \cup (-\infty, n]$ is equivalent to a single left-side interval $(-\infty, max(a, b, \ldots n)]$. The same applies for one or more right-side intervals being equivalent to $[min(a, b, \ldots n), \infty)$. Therefore, this hypothesis class is of the form $(-\infty, a] \cup [b, \infty)$.

Since $\{(-\infty, a]$ or $[b, \infty)\} \subset \{(-\infty, a] \cup [b, \infty)\}$, we know this class of hypothesis to be capable of shattering 2 points. But once again, three points $p_0, p_1, p_2, \mid p_0 < p_1 < p_2, p_0 \in \ominus, p_1 \in \oplus, p_2 \in \ominus$ cannot be shattered with this class of hypothesis. Therefore, for this class: $VC_{dim} = 2$

1.d $[a, b] \cup [c, d]$

This class of hypothesis can shatter four points due to the following:

a Any four positives can be correctly classified by a single interval as can any labeling with a single positive.

b Any two positives and two negatives can be classified with two intervals, given that a single interval is assigned to each positive.

c Labeling three positives and one negative will always yield at most two groups of contiguous positive labels, each of which can be contained in one of the two intervals.

However, if we have five points $p_0, p_1, p_2, p_3, p_4, | \; p_0 < p_1 < p_2 < p_3 < p_4, p_0 \in \oplus, p_1 \in \ominus, p_2 \in \oplus, p_3 \in \ominus, p_4 \in \oplus$ cannot be shattered with this class of hypothesis. Therefore, for this class: $VC_{dim} = 4$

1.e Unions of k intervals

By induction:

Base step: One interval, $k = 1, h = [a, b]$, and two points, $p_0, p_1 \mid p_0 < p_1, p_0$:

point	**label**	**h**
p_0	$\ominus$	$[a, b], b < p_0$
p_1	$\ominus$	
p_0	$\ominus$	$[a, b], p_0 < a < p_1 < b$
p_1	$\oplus$	
p_0	$\oplus$	$[a, b], a < p_0 < b < p_1$
p_1	$\ominus$	
p_0	$\oplus$	$[a, b], a < p_0, p_1 < b$
p_1	$\oplus$	

We increase the set to three points with the following labels $p_0, p_1, p_2, | \; p_0 < p_1 < p_2, p_0 \in \oplus, p_1 \in \ominus, p_2 \in \oplus$, it cannot be shattered Therefore, for the base step $VC_{dim} = 2 = 2k$.

Now suppose that for the union of k intervals, we can shatter $2k$ points, then we need to prove that with $k + 1$ intervals we are able to shatter $2(k + 1)$.

Firstly we note that the most *difficult* configuration to classify would be an alternation of $\oplus$ and $\ominus$ points, since it would require using each one of the k intervals to classify a single point; any other configuration would require less than k intervals and we would have some *leftover* intervals to be consumed in classifying newly inserted points.

Inductive step: We add points p_{2k}, p_{2k+1}, with no inequality constraints, to the $2k$ points shattered with k intervals. Without loss of generality, we suppose the previous points to be in an alternating configuration of labels as we mentioned above. We can contemplate three possible scenarios for the added points:

i $p_{2k}, p_{2k+1} \in \ominus$

ii $p_{2k} \in \oplus, p_{2k+1} \in \ominus$[1]

iii $p_{2k}, p_{2k+1} \in \oplus$

case i

Since the previous $2k$ points could be shattered and there are no two contiguous $\oplus$ labels in the previous set of $2k$ points, introducing two $\ominus$ labels anywhere will not disrupt prior labeling if the intervals capturing the adjacent $\oplus$ points are adjusted accordingly.

[1] Equivalent to $p_{2k+1} \in \oplus, p_{2k} \in \ominus$

case ii

As above, the $\ominus$ point will not disrupt prior labeling. The $\oplus$ point will either fall beside another $\oplus$ point where it can be included in the interval[2] capturing the adjacent $\oplus$, or at either end of the set, besides an $\ominus$ point, in which case the $k+1^{th}$ interval will correctly classify it.

case iii

If the previous $2k$ points are labeled with alternating $\ominus$ and $\oplus$, then one end of the set will have $\ominus$ and the other $\oplus$. Thus on inserting points p_{2k} and p_{2k+1} one of them will necessarily fall beside another $\oplus$ and, in the worst case, the other point could be placed at the end of the interval on the end with the $\ominus$, in which case the $k+1^{th}$ interval would correctly classify it.

Thus $k+1$ intervals shatter $2(k+1)$ points. With the addition of three points in the inductive step[3], however, with $2(k+1)+1$ points and the following configuration $\oplus, \ominus, \ldots \oplus$ we would not be able to shatter the set of points with $k+1$ intervals.

Thus the inductive step holds.

Then, for this class with k intervals, $VC_{dim} = 2k$.

Question 2: KL Divergence

2.a $KL(P||Q) \geq 0, \forall P, Q$

2.b $KL(P||Q) = 0$?

2.c Max $KL(P||Q)$?

2.d $KL(P||Q) = KL(Q||P)$? Justify

2.e Prove $KL(P(Y,X)||Q(Y,X)) = KL(P(X)||Q(X)) + KL(P(Y|X)||Q(Y|X))$

2.f Prove $\arg\min_\Theta KL(\hat{P}||P) = \arg\max_\Theta \sum_{i=1}^{m} logP_\Theta(x_i)$

Question 3: K-means

[2] Once the bounds of said interval have been adjusted

[3] For a grand-total of $1+2(k+1)$