ZHAOYANG XU + CHM564

December 17, 2019

Contents

1 Airline Revisited

Let

$$A = \{10000 \ passenagers \ with \ 95\% \ show \ up\}$$
$$B = \{100 \ passenagers \ show \ up \ for \ 99 \ seats\}$$

Now we want to calculate

$$P(A \ and \ B) = P(A)P(B)$$

By Hoeffding's Inequality:

$$\begin{aligned} P(\sum_{i=1}^{10000} X_i \geq 9500) &= P(\sum_{i=1}^{10000} X_i - E[\sum_{i=1}^{10000} X_i] \geq 9500 - E[\sum_{i=1}^{10000} X_i]) \\ &\leq e^{\frac{-2\epsilon^2}{\sum_{i=1}^{10000}(1-0)^2}} \\ &= e^{\frac{-2(9500-10000p))^2}{10000}} \end{aligned}$$

$$\begin{aligned} P(\sum_{i=1}^{10000} X_i \leq 9500) &= P(\sum_{i=1}^{10000} X_i - E[\sum_{i=1}^{10000} X_i] \leq 9500 - E[\sum_{i=1}^{10000} X_i]) \\ &\leq e^{\frac{-2\epsilon^2}{\sum_{i=1}^{10000}(1-0)^2}} \\ &= e^{\frac{-2(10000p-9500))^2}{10000}} \end{aligned}$$

$$P(A) = P(\sum_{i=1}^{10000} X_i = 9500) \leq e^{\frac{-2(10000p-9500))^2}{10000}}$$

Then we calculate the P(B)

$$P(B) = P(\sum_{i=1}^{100} X_i = 100) = p^{100}$$

$$P(A)P(B) = p^{100} e^{\frac{-2(10000p-9500))^2}{10000}}$$

Then,we calculate partial Derivative of p.

$$
\begin{aligned}
P(AB)' &= 100p^{99}e^{\frac{-2(10000p-9500))^2}{10000}} + e^{\frac{-2(10000p-9500))^2}{10000}}p^{100} * (-400)(100p-95) \\
&= 100p^{99}e^{\frac{-2(10000p-9500))^2}{10000}}(1-p(400p-385)) = 0
\end{aligned}
$$

Because$p \in [0,1]$,so we get

$$p = 0.95 \qquad\qquad P(AB) = 0.0059$$

2 The Growth Function and the VC-Dimension

1. The Growth Function of H is the maximal number of dichotomies it can generate on n points

   $$m_H(n) = \max_{x_1,x_2,...,x_n} |H(x_1,x_2,...,x_n)|$$

   Let $2^n < M$
   We can get the maximal number of dichotomies, if a set of points $x_1,x_2,...,x_n$ is shattered by H. If not, dichotmies will $< 2^n$.
   So, $m_H(n) \leq 2^n$

   Let $M < 2^n$
   Dichotomies will $\leq M$.And if a set of points $x_1,x_2,...,x_{log_2 M}$ is shattered by H, the $m_H(n) = M$.
   So,$m_H(n) \leq M$

   Then we can get $m_H(n) \leq min\{M, 2_n\}$

2. By definition

   $$d_{vc}(H) = max\{n|m_H(n) = 2^n\}$$

   From the question 2.1
   If $M > 2^n$ $\qquad$ $d_{vc}(H) \leq n$
   If $M \leq 2^n$ $\qquad$ $d_{vc}(H) \leq log_2 M$

3. If$|H| = 2$, the maximal dichotonies will be 2. There will be an incomex_1 is shattered by H, so

   $$d_{vc}(H) \geq 1$$

   Then if $d_{vc}(H) = 2$, there will 2 points be shattered. So, the dichotonies are at least 4, $|H|$ at least 4. However, $|H| \leq 4$, so

   $$d_{vc}(H) \leq 1$$

Now, we can get $d_{vc}(H) = 1$

4. Assume, we have a set of points $\{x_1, x_2, ..., x_{2n}\}$

   If $2n \geq d_{vc} \geq n$, points $\{x_1, x_2, ..., x_n\}$ still can be shattered.
   So,

$$
\begin{aligned}
&m_H(2n) = m_H(d_{vc}) = 2^{d_{vc}} \leq 2^{2n} \\
&m_H(n)^2 = 2^{n2} = 2^{2n} \\
&m_H(2n) \leq m_H(n)^2
\end{aligned}
$$

   If $d_{vc} < n$,

$$
\begin{aligned}
&m_H(2n) = m_H(d_{vc}) = 2^{d_{vc}} \\
&m_H(n)^2 = m_H(d_{vc})^2 = 2^{2d_{vc}} \\
&m_H(2n) < m_H(n)^2
\end{aligned}
$$

   Then,we can get $m_H(2n) \leq m_H(n)^2$.

5. By Theorem 3.16

$$
\begin{aligned}
\frac{8ln(\frac{2((2n)^d+1)}{\delta})}{n} &\leq 1 \\
ln(\frac{2((2n)^d+1)}{\delta}) &\leq \frac{n}{8} \\
e^{ln(\frac{2((2n)^d+1)}{\delta})} &\leq e^{\frac{n}{8}} \\
\frac{2((2n)^d+1)}{\delta} &\leq e^{\frac{n}{8}} < e^n \\
2*(2n)^d + 2 &< \delta e^n \\
2*(2n)^d &< \delta e^n < e^n \\
(2n)^d &< e^n \\
ln(2n)^d &< ln(e^n) \\
dln(2n) &< nln(e) = n
\end{aligned}
$$

   So, we can get $d < n$

6. The smaller the bound, the tighter it will be. So, we compera two generalization bounds by subtraction. (Cause we just want to know its positive or negative, we will remove some terms during process)

$$
\begin{aligned}
\sqrt{\frac{8ln(\frac{2((2n)^{d_{vc}}+1)}{\delta})}{n}} - \sqrt{\frac{ln\frac{M}{\delta}}{2n}} &\Rightarrow \frac{8ln(\frac{2((2n)^{d_{vc}}+1)}{\delta})}{n} - \frac{ln\frac{M}{\delta}}{2n} \\
&= \frac{16ln(\frac{2((2n)^{d_{vc}}+1)}{\delta})}{2n} - \frac{ln\frac{M}{\delta}}{2n} \\
&= \frac{16ln(\frac{2((2n)^{d_{vc}}+1)}{\delta}) - ln\frac{M}{\delta}}{2n} \\
&\Rightarrow 16ln(\frac{2((2n)^{d_{vc}}+1)}{\delta}) - ln\frac{M}{\delta} \\
&= 15ln(\frac{2((2n)^{d_{vc}}+1)}{\delta}) + ln(\frac{2((2n)^{d_{vc}}+1)}{\delta}) - ln\frac{M}{\delta} \\
&= 15ln(\frac{2((2n)^{d_{vc}}+1)}{\delta}) + ln(\frac{2((2n)^{d_{vc}}+1)}{\delta} * \frac{\delta}{M}) \\
&= 15ln(\frac{2((2n)^{d_{vc}}+1)}{\delta}) + ln(\frac{2((2n)^{d_{vc}}+1)}{M})
\end{aligned}
$$

Then, we pay attention to the latter term

$$
\begin{aligned}
ln(\frac{2((2n)^{d_{vc}}+1)}{M}) &\Rightarrow \frac{2((2n)^{d_{vc}}+1)}{M} \\
&= \frac{2^{d_{vc}+1}n^{d_{vc}}+2}{M}
\end{aligned}
$$

We can find the result is positive, so Therom 3.2 is tighter than Therom 3.16.

7. By VC generalization bound,

$$P(\exists h \in H : L(h) \geq \hat{L}(h,S) + \sqrt{\frac{8ln(\frac{2((2n)^{d}+1)}{\delta})}{n}}) \leq \delta$$

We need to bound loss less than 0.01 with 99% confidence. So

$$\delta = 1 - 99\% = 1\% = 0.01 = 10^{-2}$$

then, we will calculate the loss

$$\sqrt{\frac{8ln(\frac{2((2n)^d+1)}{\delta})}{n}} \leq 0.01 = 10^{-2}$$
$$\frac{8ln(\frac{2((2n)^d+1)}{\delta})}{n} \leq 10^{-4}$$
$$8ln(\frac{2((2n)^d+1)}{\delta}) \leq 10^{-4}n$$
$$ln(\frac{2((2n)^d+1)}{\delta}) \leq \frac{10^{-4}}{8}n$$
$$ln(2((2n)^d+1)*10^2) \leq \frac{10^{-4}}{8}n$$

Let $f(n) = ln(2((2n)^d+1)*10^2), \quad g(n) = \frac{10^{-4}}{8}n$, then we use python to find n, which make $f(x) \leq g(x)$. We can get results $n \geq 1.5609626 * 10^7$.(See code in code.zip)

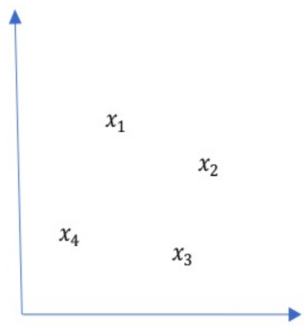
8. For a set of points $\{x_1, x_2, x_3\}$, and three points are not on a straight line. And we assume the distance between the x_i, x_j is d_{ij}.

   First, its easily to get outcomes $(+,+,+);(-,-,-)$. Just need h_+ enough big or enough small.

   Then, we use $x_i(i \in 1,2,3)$ as the centers of circles; radius $r \leq dij(i,j \in 1,2,3)$, we will get outcomes $(+,-,-)(-,+,-)(-,-,+)$.

   We connect any two points, the line splits the space into two. We make c in the space that don't have the other point.Then, we can have the outcomes$(+,+,-);(-,+,+);(+,-,+)$ Finally, we get $8 = 2^3$ dichotomies. So $d_{vc} \geq 3$.

9. For a set of points $\{x_1, x_2, x_3, x_4\}$,we assume these points are arranged clockwise,like:



   Use the method in 2.8, we can easily get outcomes by h_+:

$$(+,+,+,+)\ (+,+,+,-)\ (+,+,-,+)\ (+,-,+,+)\ (-,+,+,+)$$
$$(+,-,-,-)\ (-,+,-,-)\ (-,-,+,-)\ (-,-,-,+)\ (-,-,-,-)$$
$$(+,+,-,-)\ (-,+,+,-)\ (-,-,+,+)\ (+,-,-,+)$$

Then we choose $min\{d_{13}, d_{24}\}$, by h_+ we can get $(+,-,+,-)$ and by h_- we can get another $(-,+,-,+)$. So, we get outcomes $16 = 2^4$, $d_{vc} \geq 4$.

3 SVMs

1. For mean and variance of every feature in the transformed are 0 and 1. We define the f_{norm}

$$f_{norm} = \frac{x_i - x_{i(mean)}}{\sqrt{x_{i(var)}}}$$

We calculate the mean and variance of data of train, we can get
Train mean:

$1.55960388e+02$	$2.04821194e+02$	$1.15058622e+02$	$5.99785714e-03$
$4.28877551e-05$	$3.20418367e-03$	$3.31540816e-03$	$9.61295918e-03$
$2.77400000e-02$	$2.62408163e-01$	$1.46761224e-02$	$1.66144898e-02$
$2.19880612e-02$	$4.40281633e-02$	$2.26390816e-02$	$2.20007041e+01$
$4.94819602e-01$	$7.15689765e-01$	$5.76372753e+00$	$2.14795724e-01$
$2.36576287e+00$	$1.99708816e-01$		

Train variance:

$1.96280920e+03$	$9.63381571e+03$	$2.09357394e+03$	$1.56323454e-05$
$9.05262911e-10$	$5.59068556e-06$	$5.17943912e-06$	$5.03013168e-05$
$2.52776890e-04$	$2.64697314e-02$	$7.48602666e-05$	$1.02539466e-04$
$1.76683701e-04$	$6.73690070e-04$	$8.86782604e-04$	$1.65097280e+01$
$1.03107075e-02$	$3.11595818e-03$	$1.06174931e+00$	$5.74533305e-03$
$1.36459832e-01$	$6.65889223e-03$		

Then we the above data to compute the mean and variance of the transformed test data.
Transformed Test mean:

-0.07857931	-0.15804162	0.05562311	0.11318387	0.07157377
0.08691489	0.11567239	0.08701553	0.24898214	0.24518734
0.2295662	0.25089051	0.31660826	0.22960283	0.14905702
-0.05676346	0.07356766	0.08676698	0.15477245	0.31069455
0.08741643	0.1685766			

Transformed Test variance:

0.73218508	0.71491336	0.79759033	1.99040214	1.66604029
2.13673681	1.92226228	2.13767335	1.77195651	1.82895633
1.7173149	1.77783879	2.19022855	1.7174543	2.66297002
1.36090146	1.08263293	0.95130846	1.21651005	1.36280271
1.13351689	1.41470112			

2. We want to make a model and use this model to predict the Test data. So, our first aim to find the best parameters C, γ.We set the scope of the parameter grid search $C; \gamma \in \{0.001, 0.01, 0.1, 1, 10, 100\}$. Then we will have 36 combinations.
   For each combinations ,we do cross-validation. We select random samples from Traindata to train and other for validation and this step will repeat 5 times. At the end, we will get a score for this parameters' combinations. After try all the combinations, we choose the parameters which get the highest score.
   In my model, Train data and Transformed train data (form Question 3.1) are all used to compute the best parameters.
   For train data are $C = 10 \quad \gamma = 0.01$
   For transformed train data are $C = 1 \quad \gamma = 0.01$ (I choose the latter one to compute the loss function.)
   The test error : 22
   The train erro : 12