CSDS 440: Assignment 8

Shaochen (Henry) ZHONG, sxz517

Mingyang TIE, mxt497

Due and submitted on 10/30/2020
Fall 2020, Dr. Ray

Problem 33

The decision surface boundary is defined as $wx+b=0$. When $|c_1| \neq |c_2|$, the decision surface will be closer to either the plus or minus planes. If $|c_1| > |c_2|$, the decision surface boundary is closer to minus planes, because when $|c_1| > |c_2|$, $c_1 + c_2 > 0$, so $wx + b > 0$. If $|c_1| < |c2|$, we know that $c_1 + c_2 < 0$, therefore the decision surface boundary is closer to plus plane as $wx + b < 0$. If we have $|c_1| = |c_2|$, $wx + b = 0$, it means the resulting decision surface is halfway between plus and minus plane, which is the general SVM case.

When choosing the decision surface boundary, the preference it is really depending on the dataset we have and the task we are doing. For example, if the distribution of the dataset we have shows much less example with negative label, we may want to let the decision surface boundary to be closer the plus plane so that more "freedom" is left for negative-labeled outputs.

However, if the purposed task is to identify fire in building, a decision boundary which is closer to minus plane will be preferred as it more examples will be classifed as positive – which is a preferred bias in this task as the cost of having a couple more false positives is much more tolerable than having some false negatives.

If the task and dataset is generally neutral, then we may always redefine the decision boundary to be $wx + b = \frac{1}{2}(c_1 + c_2)$ to have an equal distance toward both the plus and minus plane.

Problem 34

With $wx + b = 0$, we know that w is perpendicular decision plane as we may select two points on w, we have $w(u - v) = wu - wv = -b - (-b) = 0$. Known that the plus and minus plane are parallel to the decision plane, w is perpendicular to both the plus and minus plane.

Say we have x^+ on plus plane, let x^- to be a point on minus plane with shortest distance to x^+. We know that there must be $x^+ - x^- = \lambda w$ since the shortest distance between two paralleled planes are on their norm. Thus, we have:

$$
\begin{aligned}
1 &= w(x^- + \lambda w) + b \\
&= wx^- + \lambda w \cdot w + b \\
\lambda ||w||^2 &= 2
\end{aligned}
$$

Since we know the margin is defined as $M = \lambda ||w||$, we have $M = \frac{2}{||w||}$ which is independent of b.

Problem 35

We denotes $F = e_A - e_B$ to represent the difference between the error rate of the two classifers. We know that $E(F) = e_1 - e_B$ and the variance will be $V(F) = \frac{e_A(1-e_A)+e_B(1-e_B)}{n}$ according to the property of binomial distribution.

As we may model our calculation as a Gaussian distribution, a 95% CI is around 1.96σ. We want to make 0 will not in this interval, thus $0.1 - 1.96\sigma > 0 \implies 0.1 > 1.96\sigma$. Now by substitute the $V(F)$ in, we have:

$$
\begin{aligned}
0.1 &> 1.96\frac{e_A(1-e_A)+e_B(1-e_B)}{n} \\
&> 1.96\frac{\sqrt{e_A(1-e_A)+e_B(1-e_B)}}{\sqrt{n}} \\
\sqrt{n} &> \frac{1.96}{0.1}\sqrt{e_A(1-e_A)+e_B(1-e_B)} \\
\implies n &> 384.16(e_A(1-e_A)+e_B(1-e_B))
\end{aligned}
$$

Problem 36

Known that the 95% CI are $[x_B, y_B]$ and $[x_N, y_N]$ respectively for two datasets, we may deduce the error rates to be $e_B = \frac{x_B+y_B}{2}$ and $e_N = \frac{x_N+y_N}{2}$ for their respective dataset.

Assume Prof. Bob's dataset has n_B examples and Prof. Nan has n_N examples, Prof. Scoop may simply derive his version of error rate to be a combination of the both, which is $e_S = \frac{n_B e_B + n_N e_N}{n_B + n_N}$. This implies $n_S = n_B + n_N$, we may get this n_S (and also n_B, n_N) by doing:

$$
\begin{aligned}
x_B &= e_B - 2\sigma_B \\
2\sqrt{\frac{e_B(1-e_B)}{n_B}} &= e_B - x_B \\
\frac{4e_B(1-e_B)}{n_B} &= (e_B - x_B)^2 \\
\implies n_B &= \frac{4e_B(1-e_B)}{(e_B-x_B)^2}
\end{aligned}
$$

By doing the same calculation for n_N we have:

$$
\begin{aligned}
n_N &= \frac{4e_N(1-e_N)}{(e_N-x_N)^2} \\
\implies n_S &= \frac{4e_B(1-e_B)}{(e_B-x_B)^2} + \frac{4e_N(1-e_N)}{(e_N-x_N)^2}
\end{aligned}
$$

Since e_B, n_B, e_N, n_N, n_S are now all known, we can calculate e_S accordingly. Now, assume Prof. Scoop will land with a 95% CI of $[x_S, y_S]$, then there must be:

$$
\begin{aligned}
x_S &= e_S - 2\sigma_S \\
&= e_S - 2\sqrt{\frac{e_S(1-e_S)}{n_S}} \\
y_S &= e_S + 2\sigma_S \\
&= e_S + 2\sqrt{\frac{e_S(1-e_S)}{n_S}}
\end{aligned}
$$

The above two equations suggest we may derive x_S and y_S with just e_S and n_S, which are two known value to us and also to Prof. Scoop. So Prof. Scoop – while not doing any experiment – may derive his 95% CI base on Prof. Bob and Prof. Nan's results at the cost of his academic integrity.