COMP0078 CW 2 3(e)

January 17, 2022

Question 3(e)

Includes `.nb` file in appendix

To start on this question, we make the following definition.

Definition 1. Let the distance between two vectors $\mathbf{x}, \mathbf{z} \in \{-1, 1\}^n$ be

$$\mathrm{d}(\mathbf{x}, \mathbf{z}) \equiv \frac{\sqrt{(\mathbf{x}-\mathbf{z})^2}}{2}, \tag{1}$$

This is just half the Euclidean distance, and so consistent with the `1NN` algorithm, but our definition is slightly easier to use in the current context. In particular, $\mathrm{d}(\mathbf{x}, \mathbf{z})$ counts the coordinates in which $\mathbf{x}$ and $\mathbf{z}$ differ. Before the main theorem, we will prove three lemmas.

Lemma 1. *For 'just a little bit' with n coordinates, if a uniformly randomly drawn vector, $\mathbf{x}$, is at a distance d from a test vector, $\mathbf{z}$, then:*

a. $\mathrm{P}\left(\mathrm{d}(\mathbf{x}, \mathbf{z}) = d\right) = \mathrm{Binom}\left(d \middle| n, \frac{1}{2}\right)$, *where* Binom *is the binomial distribution.*

b. $\mathrm{P}\left(\mathbf{x}_1 \neq \mathbf{z}_1 \middle| \mathrm{d}(\mathbf{x}, \mathbf{z}) = d\right) = \frac{d}{n}$.

Proof. By counting the number of ways to vary $\mathbf{z}$ in exactly d coordinates, we see that there are $\binom{n}{d}$ vectors in $\{-1, 1\}^n$ at distance d from $\mathbf{z}$, out of the total of $2^n = |\{-1, 1\}^n|$, Of these, we can similarly see that $\binom{n-1}{d-1}$ vectors at distance d disagree with $\mathbf{z}$ in the first coordinate, because then exactly $d-1$ coordinates other than the first must vary from those of $\mathbf{z}$. Hence, for a uniformly randomly drawn $\mathbf{x} \in \{-1, 1\}^n$,

$$\mathrm{P}\left(\mathrm{d}(\mathbf{x}, \mathbf{z}) = d\right) = \mathrm{Binom}\left(d \middle| n, \frac{1}{2}\right), \tag{2}$$

and, again using our counts of vectors at distance d,

$$\mathrm{P}\left(\mathbf{x}_1 \neq \mathbf{z}_1 \middle| \mathrm{d}(\mathbf{x}, \mathbf{z}) = d\right) = \frac{\binom{n-1}{d-1}}{\binom{n}{d}} = \frac{(n-1)!}{(d-1)!\,(n-1-[d-1])!} \frac{d!\,(n-d)!}{n!} = \frac{d}{n}. \tag{3}$$

□

The following lemma and proof are adapted to our context from a theorem in MIT lectures available online.[1]

Lemma 2. *Let $n > 0$ be an integer, and $B(n)$ be the random variable defined by* $\mathrm{Binom}(n, 1/2)$, *and define the* centred binomial random variable $X(n) \equiv B(n) - n/2$. *If $X_i(n), i = 1, ..., m$ are i.i.d. centred binomial random variables, then*

$$\frac{1}{n}\mathrm{E}\left[\max_{i=1}^{m} X_i(n)\right] \leq \frac{\log m^{1/n}}{\cosh^{-1}(m^{1/n})}. \tag{4}$$

Proof. Recall, or easily show, that

$$\mathrm{E}\left[e^{sB(n)}\right] = \left(\frac{1+e^s}{2}\right)^n, \tag{5}$$

and, hence,

$$\mathrm{E}\left[e^{sX(n)}\right] = \left(\frac{1+e^s}{2}\right)^n e^{-sn/2} = \cosh^n\left(\frac{s}{2}\right). \tag{6}$$

[1] Theorem 1.14 in Chapter 1 of MIT lectures on High–Dimensional Statistics, Spring 2015. Link to version accessed on 18 Dec.

For any $s > 0$, we then have

$$\begin{aligned}
\frac{1}{n}\mathrm{E}\left[\max_{i=1}^{m} X_i(n)\right] &= \frac{1}{ns}\mathrm{E}\left[\log e^{s \max_{i=1}^{m} X_i(n)}\right] \\
&\le \frac{1}{ns}\log \mathrm{E}\left[e^{s \max_{i=1}^{m} X_i(n)}\right] \quad \text{by Jensen's inequality} \\
&= \frac{1}{ns}\log \mathrm{E}\left[\max_{i=1}^{m} e^{sX_i(n)}\right] \\
&\le \frac{1}{ns}\log \sum_{i=1}^{m} \mathrm{E}\left[e^{sX_i(n)}\right] \\
&= \frac{1}{ns}\log\left(m\mathrm{E}\left[e^{sX_i(n)}\right]\right) \\
&= \frac{\log\left(m^{1/n}\right) + \log\cosh\left(\frac{s}{2}\right)}{s}.
\end{aligned} \tag{7}$$

Setting

$$s = 2\cosh^{-1}(m^{1/n}), \tag{8}$$

we find

$$\frac{1}{n}\mathrm{E}\left[\max_{i=1}^{m} X_i(n)\right] \le \frac{\log m^{1/n}}{\cosh^{-1}(m^{1/n})}, \tag{9}$$

which is the required result. □

Lemma 3. *Let*

$$g(x) = \frac{x}{\cosh^{-1}(e^x)}, \tag{10}$$

for $x > 0$. Then $g(x)$ is increasing as a function of x.

Proof. Let $h(x) = g(\log(x))$, defined for $x > 1$. From, for example, the *Mathematica* notebook `part_3e.nb`, we have that

$$x\cosh^{-1}(x)\, h'(x) = 1 - \frac{x}{\sqrt{x^2-1}}\frac{\log x}{\cosh^{-1}(x)}, \tag{11}$$

where the dash denotes the derivative. Set

$$q(x) = \frac{x\log(x)}{\sqrt{x^2-1}} - \cosh^{-1}(x), \tag{12}$$

and find, as in `part_3e.nb`,

$$q'(x) = -\frac{\log(x)}{(x^2-1)^{3/2}} < 0 \quad \text{and} \quad \lim_{x\to 1} q(x) = 0, \tag{13}$$

which implies $q(x) < 0$ for $x > 1$, and therefore, from eq. (11), that $x\cosh^{-1}(x)\, h'(x) > 0$, Noting that $x\cosh^{-1}(x) > 0$ for $x > 1$, we have that $h(x)$ is increasing in x for $x > 1$. Hence, as $\log x$ is increasing in x, our result follows. □

We can now prove the theorem.

Theorem. *In the 'just a little bit' problem for vectors in $\{-1, 1\}^n$, if m is the sample complexity, that is the minimum size of training set such that $P(\text{failure}) \le 0.1$ with a uniformly randomly drawn training set of size m and uniformly randomly drawn test vector, then*

$$m \in \Omega(\exp(0.36n)). \tag{14}$$

Proof. If we let d be the `1NN` distance from the test vector to the training set, for i.i.d. binomial random variables, $B_i(n)$ and centred binomial random variables, $X_i(n)$,

$$\begin{aligned}
\mathrm{P}\left(\text{failure}\right) &= \sum_{d=0}^{n} \mathrm{P}\left(d\right)\, \mathrm{P}\left(\text{failure}|d\right) \\
&= \sum_{d=0}^{n} \mathrm{P}\left(\min_{i=1}^{m} B_i(n) = d\right) \frac{d}{n} \quad \text{from lemma 1} \\
&= \frac{1}{n}\mathrm{E}\left[\min_{i=1}^{m} B_i(n)\right] \\
&= \frac{1}{n}\mathrm{E}\left[\min_{i=1}^{m} \left(X_i(n) + \frac{n}{2}\right)\right] \\
&= \frac{1}{2} - \frac{1}{n}\mathrm{E}\left[\max_{i=1}^{m} X_i(n)\right],
\end{aligned} \tag{15}$$

the last line following from the symmetry of each random variable $X_i(n)$ about zero. Applying the sample complexity condition, we have $\mathrm{P}\left(\text{failure}\right) = 0.1$, and find

$$\frac{1}{n}\mathrm{E}\left[\max_{i=1}^{m} X_i(n)\right] = \frac{2}{5}. \tag{16}$$

Lemma 2 then gives

$$\frac{2}{5} \le \frac{\log m^{1/n}}{\cosh^{-1}(m^{1/n})} \equiv g(\log m^{1/n}), \tag{17}$$

where g is the function from lemma 3. The equation $g(x) = 2/5$ can be numerically solved, as done in `part_3e.nb`, to give $x = 0.360\ldots$. Because lemma 3 says that $g(x)$ is increasing in x, we have

$$\log m^{1/n} \ge 0.360\ldots, \tag{18}$$

implying that

$$m \ge \exp(0.36n) \tag{19}$$

which gives our result. ☐

To conclude we show, numerically, that to two significant figures, 0.36 is the best exponent in eq. (19) that we can get from the bound of eq. (7) in lemma 2. Suppose that, in that lemma, instead of setting s by the expression in eq. (8) we had chosen the value of s that minimised the bound, which would be a function

$$b(a) \equiv \min_{s}\left[\frac{a + \log\cosh\left(\frac{s}{2}\right)}{s}\right], \tag{20}$$

where $a \equiv \log m^{1/n}$ is the exponent we seek to optimise, consistent with $m \in \Omega(\exp(an))$. From eq. (7), we then have

$$b(\log m^{1/n}) \ge \frac{1}{n}\mathrm{E}\left[\max_{i=1}^{m} X_i(n)\right], \tag{21}$$

and we can take a similar approach to that in the theorem from eq. (17) onwards, replacing g by b, and using numerical evidence plotted in fig. 1 to see that $b(a)$ is increasing in a. An a giving $b(a) < 0.4$ is feasible for $m \in \Omega(\exp(an))$, although not optimal, because, comparing with eqs. (15) to (17) it corresponds to an m with

$$\mathrm{P}\left(\text{failure}\right) = \frac{1}{2} - b(a) > 0.1, \tag{22}$$

whereas a value of a giving $b(a) > 0.4$ corresponds to m with failure rate below 0.1 and is not feasible. We find in `part_3e.nb` that $b(0.36) = 0.396$, which, as promised, shows that 0.36 is very close to optimal. Exponents a such that $b(a) > 0.4$ are inconsistent with $m \in \Omega(\exp(an))$, because, eq. (17) with b instead of g would then be false. With this observation, fig. 1 indicates that 0.36 is, indeed, the highest possible value of a to two significant figures, with $b(0.37) = 0.401 > 2/5$.

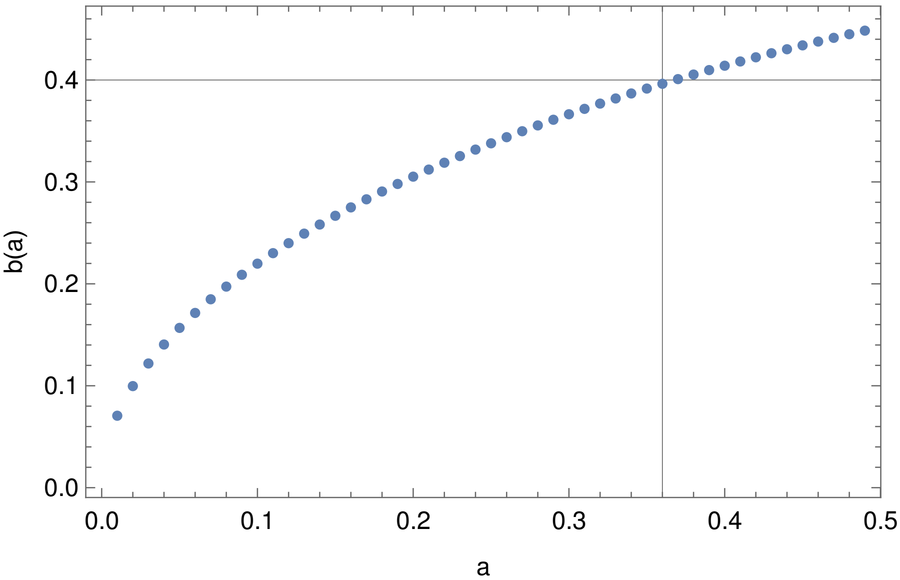


Figure 1: The function b of eq. (20). The vertical line indicates $a = 0.36$, while the horizontal line shows the optimality level of $b(a) = 2/5$. Starting from the lower values of a, increasing a become more optimal, until once $b(a) > 2/5$, they are no longer consistent with $m \in \Omega(\exp(an))$.

Appendix `part_3e.nb`

Define and analyse *g* , *h* and *q* as in Part 3e's lemma on *g(x)* increasing in *x*

In[ ]:= $g[x\_] := \frac{x}{\text{ArcCosh}[\text{Exp}[x]]}$

In[ ]:= `h[x_] := g[Log[x]]`

In[ ]:= `Assuming[x ≥ 1, Simplify[x ArcCosh[x] h'[x]]]`

Out[ ]= $1 - \frac{x\,\text{Log}[x]}{\sqrt{-1 + x^2}\;\text{ArcCosh}[x]}$

In[ ]:= $q[x\_] := x\,\text{Log}[x] \Big/ \sqrt{x^2 - 1} - \text{ArcCosh}[x]$

In[ ]:= `Assuming[x ≥ 1, Simplify[q'[x]]]`

Out[ ]= $-\frac{\text{Log}[x]}{(-1 + x^2)^{3/2}}$

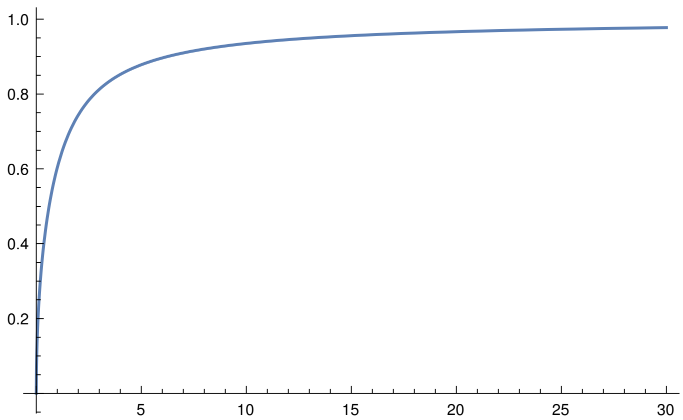
In[ ]:= `Limit[q[x], x → 1]`

Out[ ]= 0

Draw a plot showing *g(x)* increasing in *x*

In[ ]:= `Plot[g[x], {x, 0, 30}, PlotRange → Full]`

Out[ ]=



Numerically solve equation for the theorem in Part 3e

In[ ]:= `res = NSolve[g[x] == 0.4, x][[1]][[1]][[2]]`

NSolve : Warning : NSolve used FunctionExpand to transform the system . Since FunctionExpand transformation rules are only generically correct , the solution set might have been altered .

Out[ ]= 0.360214

Check the solution

In[ ]:= `N[g[res]]`

Out[ ]= 0.4

Confirm that 0.36 is very nearly the best bound from this method

In[ ]:=
```
bound[s_, a_] := (a + Log[Cosh[s/2]])/s
```

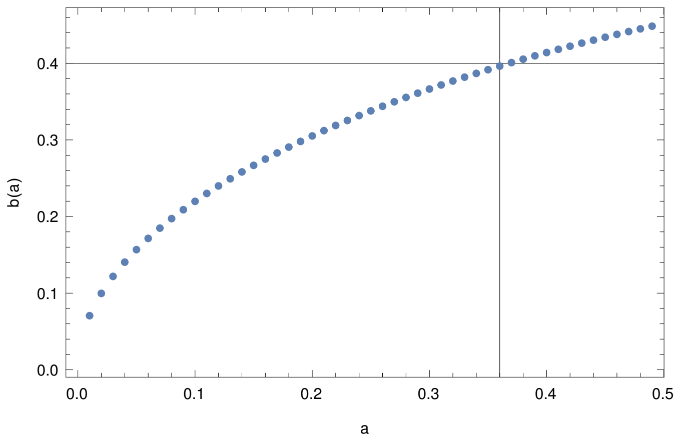
In[ ]:=
```
minbound[a_] := FindMinimum[bound[s, a], {s, a}, Method → "Newton"][[1]]
```

In[ ]:=
```
minbound[0.36]
```

Out[ ]=
```
0.396295
```

In[ ]:=
```
minbounds = {};
For[a = 0.01, a < 0.5, a = a + 0.01, AppendTo[minbounds, {a, minbound[a]}]]
```

In[ ]:=
```
ListPlot[minbounds, GridLines → {{0.36}, {0.4}}, FrameLabel → {"a", "b(a)"},
 LabelStyle → Black, GridLinesStyle → Black, Frame → True, Background → White]
```

Out[ ]=


In[ ]:=
```
For[i = 30, i ≤ Length[minbounds], i++, Print[minbounds[[i]]]]
```

{0.3 , 0.366414 }

{0.31 , 0.371697 }

{0.32 , 0.376854 }

{0.33 , 0.381888 }

{0.34 , 0.386803 }

{0.35 , 0.391605 }

{0.36 , 0.396295 }

{0.37 , 0.400879 }

{0.38 , 0.405358 }

{0.39 , 0.409736 }

{0.4 , 0.414016 }

{0.41 , 0.418199 }

{0.42 , 0.422287 }

{0.43 , 0.426284 }

{0.44 , 0.43019 }

{0.45 , 0.434008 }

{0.46 , 0.437738 }

{0.47 , 0.441382 }

{0.48 , 0.444942 }

{0.49 , 0.448418 }