6 Appendix

6.1 Split-neighborly proofs

Theorem 3.2. *If a problem is $1/\alpha$-split-neighborly and has a coherence parameter of c, for*

$$\beta = \min(c, \frac{1}{1/\alpha + 2})$$

GBS has a worst case query cost of at most $\frac{\log n}{-\log(1-\beta)}$ and GBS has an average query cost of at most $\frac{\log n}{H(\beta)}$ where $H(p)$ is the entropy of a Bernoulli(p) random variable

Proof. This theorem will follow from the next three lemmas. □

Lemma 3.1. *If a problem is $1/\alpha$-split-neighborly and has a coherence parameter of c, then for any $V \subseteq \mathcal{H}$, $|V| \leq 1$ or there exists a test $x \in \mathcal{X}$ such that*

$$\mathbb{E}_{h \in V}[h(x)] \in [\beta, 1-\beta]$$

where β is defined as above.

Proof. Fix a subset $V \subseteq \mathcal{H}$. Assume $|V| > 1$, otherwise we are done.

From the assumption, we have a coherence parameter of

$$c \geq \beta$$

From the definition, this means that there exists a probability distribution on the tests P such that for any hypothesis h,

$$\sum_{x \in X} P(x)h(x) \in [\beta, 1-\beta]$$

Since this is true for all $h \in \mathcal{H}$, this is also true for all convex combinations. Thus,

$$\mathbb{E}_{h \in V}[\sum_{x \in X} P(x)h(x)] \in [\beta, 1-\beta]$$

$$\sum_{x \in X} P(x)\mathbb{E}_{h \in V}[h(x)] \in [\beta, 1-\beta]$$

For simplicity, define the split constant $S(x) = \mathbb{E}_{h \in V}[h(x)]$. Thus,

$$\sum_{x \in X} P(x)S(x) \in [\beta, 1-\beta]$$

There are two possibilities, either there exists a test x such that

$$S(x) = \mathbb{E}_{h \in V}[h(x)] \in [\beta, 1-\beta]$$

in which case, this is the exact conclusion statement and we are done, or that there exists no test with a split constant in $[\beta, 1-\beta]$. If there exists no test with a split constant in $[\beta, 1-\beta]$ but the weighted combination is in $[\beta, 1-\beta]$, then there exists tests x and x' such that $S(x) < \beta$ and $S(x') > 1-\beta$

Since the problem is $1/\alpha$-split neighborly, there exists a graph over the tests that is strongly connected. Thus, there is a path from x to x'. Since $S(x) < \beta$ and $S(x') > 1-\beta$ and since $\forall x'' \in \mathcal{X} : S(x'') \notin [\beta, 1-\beta]$, there must be an edge (x_0, x_1) along the path where $S(x_0) < \beta$ and $S(x_1) > 1-\beta$. Thus,

$$\Pr_{h \in V}[h(x_0) = 1] = \mathbb{E}_{h \in V}[h(x_0)] < \beta$$

$$\Pr_{h \in V}[h(x_1) = 1] = \mathbb{E}_{h \in V}[h(x_1)] > 1-\beta$$

Combining these two yields,

$$\Pr_{h \in V}[h(x_0) = 0 \wedge h(x_1) = 1] > 1-2\beta$$

Recall $\Delta(x_0, x_1) = \{h \in \mathcal{H} : h(x_0) = 0, h(x_1) = 1\}$

$$\Pr_{h \in V}[h \in \Delta(x_0, x_1)] > 1-2\beta$$

$$\frac{|V \cap \Delta(x_0, x_1)|}{|V|} > 1-2\beta$$

Recall from the definition of β that $\frac{1}{1/\alpha+2} \geq \beta$. Thus

$$1-2\beta \geq 1 - 2\frac{1}{1/\alpha + 2} = \frac{1/\alpha}{1/\alpha + 2} \geq \frac{\beta}{\alpha}$$

Thus,

$$\frac{|V \cap \Delta(x_0, x_1)|}{|V|} > \frac{\beta}{\alpha}$$

For brevity, define $\Delta = \Delta(x_0, x_1)$. Since there is an edge (x_0, x_1) in the $1/\alpha$-neighborly graph, for any subset including $V \cap \Delta \subseteq \Delta$, either $|V \cap \Delta| \leq 1$ or there exists a test $\hat{x}$ such that,

$$\mathbb{E}_{h \in V \cap \Delta}[h(\hat{x})] \in [\alpha, 1-\alpha]$$

First, $|V \cap \Delta| \neq 0$, since $|V| > 1$ and $\frac{|V \cap \Delta(x_0, x_1)|}{|V|} > \frac{\beta}{\alpha}$. If $|V \cap \Delta| = 1$, then, $\frac{|V \cap \Delta(x_0,x_1)|}{|V|} > \frac{\beta}{\alpha}$ and $|V| > 1$ so $\frac{1}{2} \geq \frac{1}{|V|} > \frac{\beta}{\alpha} \geq \beta$. Since the hypotheses are identifiable, any pair of hypotheses yield a different result on some test, so we can always find a test with a split constant of at least $\frac{1}{|V|}$, and this implies the result of the theorem.

In the other case, where $|V \cap \Delta| > 1$, we have all the necessary pieces and it's just a matter of crunching the algebra.

$$\begin{aligned}
\mathbb{E}_{h \in V}[h(\hat{x})] &= \frac{\sum_{h \in V} h(\hat{x})}{|V|} \\
&\geq \frac{\sum_{h \in V \cap \Delta} h(\hat{x})}{|V|} \\
&\geq \frac{\beta}{\alpha} \frac{\sum_{h \in V \cap \Delta} h(\hat{x})}{|V \cap \Delta|} \\
&\geq \frac{\beta}{\alpha} \mathbb{E}_{h \in V \cap \Delta}[h(\hat{x})] \\
&\geq \frac{\beta}{\alpha} \alpha = \beta
\end{aligned}$$

Additionally,

$$\begin{aligned}
\mathbb{E}_{h \in V}[h(\hat{x})] &= \frac{\sum_{h \in V} h(\hat{x})}{|V|} \\
&= \frac{\sum_{h \in V \cap \Delta} h(\hat{x}) + \sum_{h \in V \setminus \Delta} h(\hat{x})}{|V|} \\
&\leq \frac{(1-\alpha)|V \cap \Delta| + \sum_{h \in V \setminus \Delta} h(\hat{x})}{|V|} \\
&\leq \frac{(1-\alpha)|V \cap \Delta| + |V| - |V \cap \Delta|}{|V|} \\
&\leq 1 - \alpha \frac{|V \cap \Delta|}{|V|} \\
&\leq 1 - \alpha \frac{\beta}{\alpha} = 1 - \beta
\end{aligned}$$

Thus, we have that

$$\mathbb{E}_{h \in V}[h(\hat{x})] \in [\beta, 1-\beta]$$

which is the conclusion of the lemma.

□

Lemma 6.1. *If, for any $V \subseteq \mathcal{H}$, $|V| \leq 1$ or there exists a test $x \in \mathcal{X}$ such that*

$$\mathbb{E}_{h \in V}[h(x)] \in [\beta, 1-\beta]$$

then GBS has a worst case query cost of at most $\frac{\log n}{\log(\frac{1}{1-\beta})}$

Proof. After m queries, there are at most $\max(1, (1-\beta)^m n)$ remaining hypotheses since greedy will choose a test with a split constant of at least β (a split with respect to the hypotheses without a prior) and will terminate when there is a single hypothesis. Thus, when $(1-\beta)^m n \leq 1$, the algorithm must have terminated. Rearranging, we see that when $m \geq \frac{\log n}{\log(\frac{1}{1-\beta})}$ the algorithm must have terminated. This means that the worst case query cost must be at most $\frac{\log n}{\log(\frac{1}{1-\beta})}$. □

Lemma 6.2. *If, for any $V \subseteq \mathcal{H}$, $|V| \leq 1$ or there exists a test $x \in \mathcal{X}$ such that*

$$\mathbb{E}_{h \in V}[h(x)] \in [\beta, 1-\beta]$$

then GBS has an average query cost of at most $\frac{\log n}{H(\beta)}$ where $H(p)$ is the entropy of a Bernoulli(p) random variable

Proof. Define $H(p)$ as the entropy of a Bernoulli random variable with probability p.

$$f(V) = \mathbb{E}[\text{average queries remaining while at subset } V] \tag{1}$$

We will prove by induction on increasing subsets that

$$f(V) \leq \frac{\log(|V|)}{H(\beta)} \tag{2}$$

Note that the base case is that $f(\{h\}) = 0$ because we are done when there is just one hypothesis left. Note that this suffices to show that the total runtime is $\log(n)/H(\beta)$ because $|V| = n$ at the beginning of the algorithm.

Let A, B be a partition of V based on a test split. Without loss of generality, let $|A| \leq |B|$, so $|A| \leq 1/2|V|$. Based on the recursive definition of cost and there is a test with a split constant of at least β (so GBS will choose a test with a split constant of at least β),

$$f(V) \leq \max_{A, B, |A|/|V| \in [\beta, 1/2]} \frac{|A|}{|V|} f(A) + \frac{|B|}{|V|} f(B) + 1$$

From the induction hypothesis,

$$\leq \max_{\dots} \frac{|A|}{|V|}\frac{\log|A|}{H(\beta)} + \frac{|B|}{|V|}\frac{\log|B|}{H(\beta)} + 1$$

$$\leq \frac{\max_{\dots} \frac{|A|}{|V|}\log|A| + \frac{|B|}{|V|}\log|B| + H(\beta)}{H(\beta)}$$

$$\leq \frac{\max_{\dots} \frac{|A|}{|V|}\log\frac{|A|}{|V|} + \frac{|B|}{|V|}\log\frac{|B|}{|V|} + H(\beta) + \log|V|}{H(\beta)}$$

$$\leq \frac{\max_{\dots} -H(\frac{|A|}{|V|}) + H(\beta) + \log|V|}{H(\beta)}$$

Note that since $|A|/|V| \in [\beta, 1/2]$ (the condition of the max), $H(\frac{|A|}{|V|}) \geq H(\beta)$. Thus, the max is non-positive, and thus,

$$f(V) \leq \frac{\log(|V|)}{H(\beta)}$$

Thus, we have proved the statement by induction and this suffices to show that the total runtime is at most $\log(n)/H(\beta)$.

□

Proposition 3.1. *If a problem is k-neighborly and has a uniform prior, then the problem is k-split-neighborly.*

Proof. In the case that $k = 1$, $|\Delta(x, x')| = 1$ so $|V| \leq 1$ so the problem is 1-split-neighborly. Assume $k > 1$. Note that any set of hypotheses must have a test that distinguishes at least one of the hypotheses (otherwise the hypotheses are the same). If two points x and x' in the k-neighborly graph have an edge between them, then $|\Delta(x, x') \cup \Delta(x', x)| \leq k$, which implies $|\Delta(x, x')| \leq k$, and thus either $|\Delta(x, x')| \leq 1$ or there is a test with a $1/k$ split constant and thus there is an edge from x to x' in the k-split-neighborly graph. By a similar argument, there is also an edge from x' to x. Since the k-neighborly graph is connected and each edge corresponds to a bidirectional edge in the k-split-neighborly graph, the k-split-neighborly graph is strongly connected and thus the problem is k-split-neighborly. □

6.2 Value of k

6.2.1 Disjunctions

For the disjunctions problem, for $m \geq 2, d \geq 2m$,

$$n = \sum_{i=1}^{m} \binom{d}{i}$$

$$k \geq \sum_{i=1}^{m} \binom{d-1}{i-1}$$

$$k \geq 1 + \sum_{i=1}^{m-1} \binom{d-1}{i}$$

$$k^2 - n \geq 1 + 2\sum_{i=1}^{m-1} \binom{d-1}{i} + (\sum_{i=1}^{m-1} \binom{d-1}{i})^2$$

$$- \sum_{i=1}^{m-1} \binom{d}{i} - \binom{d}{m}$$

Note that $2\binom{d-1}{i} \geq \binom{d}{i}$ since $i \leq m-1 \leq d/2$.

$$k^2 - n \geq 1 + (\sum_{i=1}^{m-1} \binom{d-1}{i})^2 - \binom{d}{m}$$

$$\geq \binom{d-1}{m-1}^2 - \binom{d}{m}$$

$$\geq \binom{d-1}{m-1}(\binom{d-1}{m-1} - d/m)$$

Since $m \geq 2$,

$$\geq \binom{d-1}{1} - d/2$$

$$\geq d/2 - 1$$

$$\geq m - 1$$

$$\geq 0$$

Thus, $k^2 - n \geq 0$ and so $k \geq \sqrt{n}$.

6.2.2 Monotonic CNF

Note that $n = |\mathcal{H}| = \frac{1}{l!}\binom{d}{m,m,\dots,m,d-lm}$. All of the bit strings with strictly less than l ones will be trivially connected in the k-neighborly graph, because they yield 0 on all hypotheses. However, the closest test to connect them to the rest of the graph is the bit string $1^l0^{d-l} \in \mathcal{X}$, which disagrees on $\binom{d-l}{m-1,m-1,\dots,m-1,d-lm} \leq k$ hypotheses. We examine the case where $d \geq 2ml$ and $m \geq 2$.

For the monotonic CNF formulas, recall that

$$n = |\mathcal{H}| = \frac{1}{l!}\binom{d}{m, m, \dots, m, d - lm}$$

$$k \geq \binom{d-l}{m-1, m-1, ..., m-1, d-lm}$$

For $d \geq 2ml$ and $m \geq 2$, $k \geq \sqrt{n}$.

$$\binom{d-l}{m-1, m-1, ..., m-1, d-lm} \leq k$$

and

$$n = \frac{1}{l!}\binom{d}{m, m, ..., m, d-lm}$$
$$= \frac{1}{l!}\frac{d!}{(m!)^l(d-lm)!}$$
$$= \frac{(d-l)!}{(m-1)!^l(d-lm)!}\frac{1}{m^l}\frac{d!(d-2l)!}{(d-l)!^2}\frac{(d-l)!}{l!(d-2l)!}$$
$$\leq k\frac{1}{m^l}\frac{d!(d-2l)!}{(d-l)!^2}\binom{d-l}{l}$$

Since $d \geq 2ml \geq 4l$,

$$n \leq k\frac{2^l}{m^l}\binom{d-l}{l}$$

Since $d-l \geq 2l(m-1)$ and $m \geq 2$

$$n \leq k\binom{d-l}{l(m-1)}$$
$$n \leq k\binom{d-l}{m-1, m-1, ..., m-1, d-lm}$$
$$n \leq k^2$$
$$k \geq \sqrt{n}$$

6.2.3 Discrete Linear Classifier

Recall that we are in the special case where d is divisible by 4, $b = d/4 - 1$ and there are an equal number of 1 and 0 weights ($d/2$).

All tests with fewer than $d/4$ 1's will yield a result of 0 for all hypotheses. The test with the next fewest hypotheses that yield 1 will be a test with exactly $d/4$ 1's. Thus, k is at least the number of such hypotheses that yield 1.

$$n = \binom{d}{d/2}$$
$$k \geq \binom{3d/4}{d/4}$$

For simplicity, define $c = d/4$.

$$\frac{n}{k^2} \leq \frac{\binom{4c}{2c}}{\binom{3c}{c}^2}$$

$$= \frac{(4c)!c!c!}{(3c)!(3c)!}$$

Note that we have the common Stirling's approximation,

$$\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq en^{n+1/2}e^{-n}$$

Thus,

$$\frac{n}{k^2} \leq \frac{e^3(4c)^{4c+1/2}c^{c+1/2}c^{c+1/2}e^{-6c}}{2\pi(3c)^{3c+1/2}(3c)^{3c+1/2}e^{-6c}}$$

$$= \frac{2e^3\sqrt{c}(4c)^{4c}c^cc^c}{6\pi(3c)^{3c}(3c)^{3c}}$$
$$= \frac{e^3\sqrt{c}4^{4c}}{3\pi 3^{3c}3^{3c}}$$
$$= \frac{e^3}{3\pi}\sqrt{c}(\frac{256}{729})^c$$
$$\leq 1$$

for $c \geq 1$.

Thus, for $d \geq 4$,

$$\frac{n}{k^2} \leq 1$$

$$k \geq \sqrt{n}$$

6.3 Necessity of Dependencies

6.3.1 Linear classifiers on convex polygon data pool

For arbitrary data points where the points are not the vertices of a convex polygon, the linear classifier problem is not $1/\alpha$-split-neighborly for constant α. A counter-example is shown in Figure 8.

6.3.2 Disjunctions

The linear dependence on m for the disjunctions is necessary because of the case where $d = m + 1$, and $|\mathcal{H}| = d$ (each $h \in \mathcal{H}$ lacking one variable). In this case, there are no tests with split constants of $\frac{1}{m}$, so the problem cannot be better than $(m-2)$-split-neighborly (recall coherence $c = 1/2$).

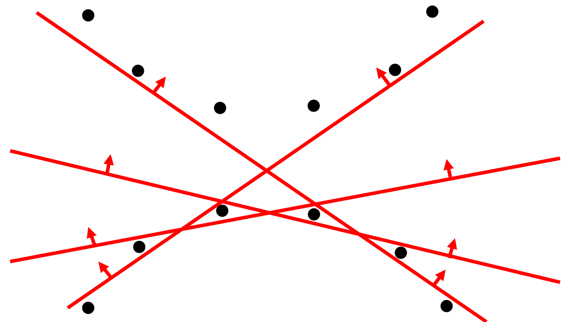

Figure 8: A counterexample that shows a non-convex data pool need not be split-neighborly. Note that we can at most split off 1 of the n hypotheses by querying one of the points from the lower half. However, the problem has coherence close to 1/2 and thus it cannot be $1/\alpha$-split-neighborly for constant α.

6.3.3 Monotonic CNF

For the monotonic CNF problem, the linear dependence on m is necessary because of the case where $l = 1$, $d = m + 1$, and $|\mathcal{H}| = d$ (each $h \in \mathcal{H}$ lacking one variable). In this case, there are no tests with split constants of $\frac{1}{m}$, so the problem cannot be better than $(m - 2)$-split-neighborly (recall coherence $c = 1/2$). Furthermore, the linear dependence on l is necessary because of the problem where $m = 1$, $d = l + 1$, and $|\mathcal{H}| = d$ (each $h \in \mathcal{H}$ lacking one variable). For this problem, there are no tests with split constants of $\frac{1}{l}$, so the problem cannot be better than $(l - 2)$-split-neighborly. Thus, although the linear dependence on m and l is necessary, it may be possible to improve the constants.

6.3.4 Object Localization

For object localization with the axis-symmetric, axis-convex set S, the dependence on d is necessary because if we use the set $S = \{je_i : |j| \leq l, 1 \leq i \leq d\}$ and consider the set of hypotheses, $\{\pm le_i : 1 \leq i \leq d\}$, the problem has no test with split constant of $\frac{1}{2d-1}$ but has coherence $c = 1/2$, so it can't be $(2d - 3)$-split-neighborly.

6.4 Monotonic CNF

Theorem 4.2. *The Conjunction of Disjunctions problem is* $(m + 1 + 3(l - 1))$*-split-neighborly.*

Proof. We prove this theorem by induction on l. First, for the base case $l = 1$.

The test graph has an edge from x to x' if $||x-x'||_1 = 1$ (the bit strings differ in one location).

Let x^+ be the value of x or x' with more 1's (and let x^- be the other one). Note that $|\Delta(x^+, x^-)| = 0$ so there is a directed edge (x^+, x^-).

For the other direction, fix a subset $V \subseteq \Delta(x^-, x^+)$. Without loss of generality, let x^+ and x^- differ in the first coordinate so $x_1^+ = 1$ and $x_1^- = 0$ and $\forall i > 1 : x_i^+ = x_i^-$.

For a proof by contradiction, the problem is not $(m + 1)$-split-neighborly so that $|V| > 1$ and there is no test x such that $\mathbb{E}_{h\in V}[h(x)] \in [q, 1 - q]$, where $q = 1/(m + 1)$.

Let

$$\mathcal{X}^+ = \{x \in \mathcal{X} : \Pr_{h\in V}[h(x) = 1] > 1 - q\},$$

$$\mathcal{X}^- = \{x \in \mathcal{X} : \Pr_{h\in V}[h(x) = 1] < q\} = \mathcal{X} - \mathcal{X}^+.$$

Let x' be the the element of $\mathcal{X}^-$ with the fewest 0's and let the 0's be at indices Z (note $1 \in Z$). If $|Z| < m$, then $h(x') = 1$ for all h since the disjunctions have m variables. But since $x' \in \mathcal{X}^-$, which is a contradiction.

Define $\{x^{(j)}\}_{j\in Z}$ to be the test resulting changing the j^{th} bit of x' to a 1. By the minimal definition of x', $\forall j \in Z : x^{(j)} \in \mathcal{X}^+$.

Suppose $|Z| > m$. Take a subset $Z' \subseteq Z$ such that $|Z'| = m + 1$. Then, from the definition of $\mathcal{X}^+$ and $\mathcal{X}^-$, $\Pr_{h\in V}[h(x') = 0 \wedge \forall j \in Z' : h(x^{(j)}) = 1] > 1 - (m + 1)q \geq 0$, which means $\Pr_{h\in V}[h \text{ includes variables } Z'] > 0$. Therefore, there is a disjunction with at least $m+1$ variables, which is a contradiction.

Thus, $|Z| = m$, so there is only one hypothesis such that $h(x') = 0$, the hypothesis with variables at Z. So $1/|V| > 1 - q$ (by definition of $\mathcal{X}^-$), which implies $|V| = 1$ since $q \leq 1/2$, which is a contradiction. Thus, by contradiction, the problem with $l = 1$ is $(m + 1)$-split-neighborly. For $l > 1$, we proceed by induction. We can define the graph as above, define $\mathcal{X}^-$ and $\mathcal{X}^+$ as above, and x' and Z as above. The same argument goes through that $|Z| = m$. Thus, $(1 - q)$ proportion of the hypotheses have a disjunction with variables at the indices Z. These hypotheses are simply another copy of the problem with $l-1$ conjunctions and $d-m$ variables. Since that problem has 1/2 coherence and is $m + 1 + 3(l - 2)$-splittable (by induction hypothesis), there exists some test with a split constant of $\frac{1}{m+1+3(l-2)+2}$ for a total split constant on the original problem of

$$(1 - q)\frac{1}{m + 1 + 3(l - 2) + 2} = \frac{1}{m + 1 + 3(l - 1)}$$

Thus, the problem is $m + 1 + 3(l - 1)$-split-neighborly by induction. $\square$

6.5 Box Object Localization

Theorem 4.3. *The object localization problem where S is a box is 4-split-neighborly.*

Notationally, refer to z_h as the integer vector for the hypothesis h and $z_{h,i}$ to be its i^{th} component.

We begin by fixing two tests x and x' such that $||x - x'||_1 = 1$. Without loss of generality, let $x' - x = e_1$ where e_1 is the 1^{st} elementary vector. Since the box is axis symmetric, there exists radii $r_i \geq 0$ such that $x - z_h \in S \leftrightarrow \forall i : |x_i - z_{h,i}| \leq r_i$. Without loss of generality, assume $x = (r_1, 0, 0, ..., 0)$ and $x' = (r_1 + 1, 0, 0, ..., 0)$. Recall $\Delta(x, x') = \{h : h(x) = 0 \wedge h(x') = 1\}$, this implies that $\Delta(x, x') = \{h : z_{h,1} = 0 \wedge \forall i > 1 : |z_{h,i}| \leq r_i\}$. We will begin by fixing a subset $V \subseteq \Delta(x, x')$. As in all the application proofs, we will start by assuming by contradiction that there is no test with a split constant in the range $[q, 1 - q]$ where $q = 1/4$. We will use this contradiction to show that the size of V is small, so that there is in fact a test with a split constant q which is a contradiction.

6.5.1 Majority Element

Fix a dimension i. Examine the tests $X_i = \{je_i : j = 0, .., 2r_i + 1\}$ and note that for $h \in V \subseteq \Delta(x, x')$, $h(je_i) = \mathbb{1}[z_{h,i} \geq j - r_i]$.

By the contradiction assumption,

$$\mathbb{E}_{h \in V}[h(je_i)] \notin [q, 1 - q]$$

$$\Pr_{h \in V}[z_{h,i} \geq j - r_i] \notin [q, 1 - q]$$

Since $\Pr_{h \in V}[z_{h,i} \geq -r_i] = 1$ and $\Pr_{h \in V}[z_{h,i} \geq r_i + 1] = 0$, there must be some integer m_i such that

$$\Pr_{h \in V}[z_{h,i} \geq m_i] > 1 - q$$

$$\Pr_{h \in V}[z_{h,i} \geq m_i + 1] < q$$

which implies that

$$\Pr_{h \in V}[z_{h,i} = m_i] > 1 - 2q$$

Define thus, there exists a vector m such that there is a $1 - 2q$ probability that an hypothesis' i^{th} component matches m.

6.5.2 Side Splits

Intuitively, we will create a sequence of tests that each remove at least half of the elements with the i^{th} component not equal to m. For each test in the sequence, the probability that the test yields 1 over the hypotheses in V must be greater that $1 - q$ so we can prove that there aren't many elements that disagree with m at any component.

Here we recursively define sets S_i, B_i, and A_i. S_i will be defined in terms of B_i and B_i will be defined in terms of S_{i-1}.

Define $S_0 = V$ and for $i > 1$, $S_i = S_{i-1} - B_i$. Noting that we could reflect the i^{th} component about m_i, without loss of generality, suppose that

$$\Pr_{h \in S_i}[z_{h,i} > m_i] \geq \Pr_{h \in S_i}[z_{h,i} < m_i]$$

Define $B_i = \{h \in S_{i-1} : z_{h,i} > m_i\}$ and $A_i = \{h \in S_{i-1} : z_{h,i} < m_i\}$

Note that $|B_i| \geq |A_i|$.

Further, there is a test $x^{(i)} = (-r_1, ..., -r_i, 0, ...0)$ such that $h(x^{(i)}) = 1 \leftrightarrow h \in S_i$ and thus by the contradiction assumption,

$$\frac{|S_i|}{|S|} \notin [q, 1 - q]$$

However, since $\Pr_{h \in V}[z_{h,i} = m_i] > 1 - 2q$, $|B_i|/|V| < 2q$. We now prove by induction that $|S_i|/|V| > 1 - q$. The base case is that $|S_1|/|V| = 1 > 1 - q$. As long as $q \leq 1/4$, since $|S_{i-1}|/|V| > 1 - q$ and $|B_i|/|V| < 2q$, $|S_i|/|S| > 1 - 3q \geq q$ (since $q = 1/4$)and thus by the contradiction assumption $|S_i|/|S| > 1 - q$.

Note that the B_i are disjoint because

$$B_i \subseteq S_i = V - B_1 - B_2 - ... - B_{i-1}$$

$$|S_d| > (1 - q)|V|$$

$$|V - \bigsqcup_{i=1}^{d} B_i| > (1 - q)|V|$$

$$|V| - \sum_{i=1}^{d} |B_i| > (1 - q)|V|$$

$$q|V| > \sum_{i=1}^{d} |B_i|$$

Define the set of elements $M' \subseteq V$ as the points with a component not equal to m. This is the union of all A_i and B_i,

$$\begin{aligned}|M'| &= |\bigcup_{i=1}^{d} A_i \cup \bigcup_{i=1}^{d} B_i| \\ &\leq \sum_{i=1}^{d} |A_i| + \sum_{i=1}^{d} |B_i| \\ &\leq 2\sum_{i=1}^{d} |B_i| \\ &< 2q|V|\end{aligned}$$

Also note that $|M'| \geq |V| - 1$ since there can only be one element that doesn't disagree with any element of m. Thus,

$$\begin{aligned}|V| - 1 &< 2q|V| \\ |V| &< \frac{1}{1-2q}\end{aligned}$$

Since $q \leq 1/3$, then this implies $|V| < 3$ so there is a test with a split of 1/3, which is a contradiction. So in a proof by contradiction, the problem is 4-split-neighborly.

6.6 Convex, axis-symmetric Shape Object Localization

Theorem 4.4. *If S is a bounded, axis-symmetric, axis-convex shape, the object localization problem is $(4d+1)$-split-neighborly.*

Proof. Let the test graph has an edge from x to x' if $||x - x'||_1 = 1$.

Fix a subset $V \subseteq \Delta(x, x')$. Without loss of generality, let $x' = 0^d$. $V \subseteq \Delta(x, x') \subseteq \{h : h(x') = 1\} = \{h : z_h - x' \in S\} = \{h : z_h \in S\}$

For a proof by contradiction, the problem is not $4d+1$-split-neighborly so that $|V| > 1$ and there is no test x such that $\mathbb{E}_{h\in V}[h(x)] \in [q, 1-q]$, where $q = 1/(4d+1)$.

Let

$$\mathcal{X}^+ = \{x \in \mathcal{X} : \Pr_{h\in V}[h(x) = 1] > 1 - q\}$$
$$\mathcal{X}^- = \{x \in \mathcal{X} : \Pr_{h\in V}[h(x) = 1] < q\} = \mathcal{X} - \mathcal{X}^+$$

Note that $x' = 0^d \in \mathcal{X}^+$ since $V \subseteq \{h : h(x') = 1\}$.

Fix a dimension i. Examine the set of tests $\{je_i : j \in \mathbb{Z}\}$. From above, $0e_i \in \mathcal{X}^+$. Further, since $V \subseteq \{h : z_h \in S\}$ and since S is bounded, there exists some $B \in \mathbb{Z}$ such that $\pm Be_i \in \mathcal{X}^-$. Thus there exists some $c_1 \leq 0, c_2 \geq 0$ such that $(c_1 - 1)e_i \in \mathcal{X}^-$, $c_1e_i \in \mathcal{X}^+$, $c_2e_i \in \mathcal{X}^+$, $(c_2 + 1)e_i \in \mathcal{X}^-$. From the definition of $\mathcal{X}^+$ and $\mathcal{X}^-$,

$$\begin{aligned}&\Pr_{h\in V}[h((c_1 - 1)e_i) = 0, h(c_1e_i) = 1, ... \\ &h(c_2e_i) = 1, h((c_2 + 1)e_i) = 0] > 1 - 4q\end{aligned}$$

Define $S_l = \{s_{-i} : s_i = l, s \in S\}$ to be the slices of S along axis i at location l. Therefore, $h(je_i) = 1 \leftrightarrow z_{h,-i} \in S_{z_{h,i}-j}$.

Note that $S_{-l} = S_l$ since the shape S is axis symmetric. Combining these three facts,

$$\begin{aligned}&\Pr_{h\in V}[z_{h,-i} \in S_{|z_{h,i}-c_1|} \cap S_{|z_{h,i}-c_2|} \setminus ... \\ &\setminus (S_{|z_{h,i}-(c_1-1)|} \cup S_{|z_{h,i}-(c_2+1)|})] > 1 - 4q\end{aligned}$$

Note that for $l' > l \geq 0$, $S_{i,l'} \subseteq S_{i,l}$ because of axis-convexity. To see this, suppose there was $t \in S_{i,l'} \setminus S_{i,l}$, then there would be three elements $s^{(-1)}, s^{(0)}, s^{(1)}$ such that $s_{-i}^{(j)} = t$ and $s_i^{(-1)} = -l', s_i^{(0)} = l, s_i^{(1)} = l'$, which would imply $s^{(-1)} \in S, s^{(0)} \notin S, s^{(1)} \in S$ which contradicts axis-convexity.

Thus, in order for the set composed of slices of S in the equation above to be non-empty,

$$|z_{h,i} - c_1|, |z_{h,i} - c_2| < |z_{h,i} - c_1 + 1|, |z_{h,i} - c_2 - 1|$$

it must be the case that $z_{h,i} = \frac{c_1+c_2}{2} \in \mathbb{Z}$ which we define to be m_i. So,

$$\Pr_{h\in V}[z_{h,i} = m_i] > 1 - 4q$$

Repeating this argument for all dimensions and combining,

$$\Pr_{h\in V}[\forall i : z_{h,i} = m_i] > 1 - 4dq$$

There is only one such element $z_h = m$ so

$$\frac{1}{|V|} > 1 - 4dq = 1 - 4d\frac{1}{4d+1} = \frac{1}{4d+1}$$

So $|V| < 4d + 1$ so there must be a split of at least q which is a contradiction.

$\square$

6.7 Discrete Binary Linear Classifiers

Theorem 4.5. *The discrete binary linear classifier problem is* $\max(16, 8r)$*-split-neighborly.*

Define $q = \min(\frac{1}{16}, \frac{1}{8r})$

Recall that for the Discrete Binary Linear Classifier case, we have hypotheses as a pair of vectors and threshold $h = (w_h, b_h) \in \{-1, 0, 1\}^d \times \mathbb{Z}$ and tests as vectors $\{0, 1\}^d$. Recall $h(x) = \mathbb{1}[w_h \cdot x > b_h]$.

From the problem setting of Discrete Binary Linear Classifiers, we know that,

$$w_h^{(+)} - b \le r(w_h^{(-)} + b) - \frac{d}{8}$$
$$w_h^{(-)} + b \le r(w_h^{(+)} - b - 1) - \frac{d}{8}$$

Recall $w^{(+)}$ is the number of positive elements of w and $w^{(-)}$ is the number of negative elements. Notationally $w_{h,i}$ refers to the i^{th} component of w_h.

6.7.1 Key Lemma and its Sufficiency

We will first state a lemma and then prove that it implies the problem stated stated.

Lemma 6.3. *Define*

$$x^{(0)} = (0, 0, ..., 0)$$
$$x^{(1)} = (1, 0, ..., 0)$$
$$H' = \{h \in H : h(x^{(0)} = 0 \wedge h(x^{(1)}) = 1 \wedge ...$$
$$\wedge w_h^{(+)} \le r w_h^{(-)} - \frac{d}{8} \wedge ...$$
$$... \wedge w_h^{(-)} \le r(w_h^{(+)} - 1) - \frac{d}{8}\}$$

For any subset $V \subset H'$*, there exists a test* x *such that* $\mathbb{E}_{h \in V}[h(x)] \in [q, 1-q]$

6.7.2 Proof of Theorem 4.5 from Lemma 6.3

We will prove Theorem 4.5 by a reduction to Lemma 6.3. To show that the problem is $1/\alpha$-split-neighborly, we need to show that for two tests with x and x' with $||x - x'||_1 = 1$ that for any subset $V \subseteq \Delta(x, x') = \{h : h(x) = 0 \wedge h(x') = 1\}$, that $|V| \le 1$ or there exists a test $\hat{x}$ such that

$$\Pr_{h \in V}[h(\hat{x}) = 1] \in [q, 1-q]$$

Note that by permuting the indices of x and x', we can make the first index the one that is different between x and x'. Additionally, for the remaining indices we can flip the 0's and 1's of the test so long as we flip the non-zero entries of w_h at that same position, and change b_h accordingly. We flip the bits so that x becomes $x^{(0)}$ and x' becomes $x^{(1)}$.

Note that $h(x^{(0)}) = 0$ implies that $0 \le b_h$. Further note that, $h(x^{(1)}) = 1$ implies that $w_{h,1} > b_h$. Thus, the only possibility is that $w_{h,1} = 1$ and $b_h = 0$.

Let T_{+-} denote the number of flips from positive to negative weights and let T_{-+} denote the number of flips from negative to positive. Then, the weights for the new (reduction) problem will be

$$w_{new}^{(+)} = w^{(+)} + T_{-+} - T_{+-}$$
$$w_{new}^{(-)} = w^{(-)} + T_{+-} - T_{-+}$$
$$0 = b_{new} = b - T_{+-} + T_{-+}$$

From the last equation, $b = T_{+-} - T_{-+}$. Thus,

$$w_{new}^{(+)} = w^{(+)} - b$$
$$w_{new}^{(-)} = w^{(-)} + b$$

Since,

$$w^{(+)} - b \le r(w^{(-)} + b) - \frac{d}{8}$$
$$w^{(-)} + b \le r(w^{(+)} - b - 1) - \frac{d}{8}$$

then,

$$w_{new}^{(+)} \le r w_{new}^{(-)} - \frac{1}{8}d$$
$$w_{new}^{(-)} \le r(w_{new}^{(+)} - 1) - \frac{1}{8}d$$

We can see that the hypothesis conditions for the original theorem imply that $\Delta(x^{(0)}, x^{(1)})$ is a subset of the hypotheses that satisfy the conditions based on $w_{new}^{(-)}$ and $w_{new}^{(+)}$ so Lemma 6.3 implies the binary linear classifier is $1/q$-split-neighborly which means $\max(16, 8r)$-split-neighborly.

6.7.3 Proof of Lemma 6.3

The remainder of this is devoted to proving Lemma 6.3

We begin by fixing a subset $V \subseteq H'$. As in all the application proofs, we will start by assuming by contradiction that there is no test with a split constant in the range $[q, 1-q]$. We will use this contradiction to show that the size of V is small.

Recall that $b_h = 0$ for all hypotheses in the reduced problem and $w_{h,1} = 1$. This follows from the fact that $h(x^{(0)}) = 0$ and $h(x^{(1)}) = 1$.

6.7.4 Majority Vector

Let e_i be an elementary vector with all entries 0 except for the i^{th} entry which is 1.

Lemma 6.4. *There exists a vector* $m \in \{-1, 0, 1\}^d$ *such that* $\forall i : m_i = 0 : \Pr_{h\in V}[w_{h,i} = m_i] \geq 1 - 2q$ *and* $\forall i : m_i \neq 0 : \Pr_{h\in V}[w_{h,i} = m_i] \geq 1 - q$

Proof. By the contradiction assumption, there isn't a test with a split constant greater than q,

$$\Pr_{h\in V}[w_h \cdot e_i > b_h] \notin [q, 1-q]$$

$$\Pr_{h\in V}[w_{h,i} > 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[w_{s,i} = 1] \notin [q, 1-q]$$

Also, by the contradiction assumption,

$$\Pr_{h\in V}[w_h \cdot (e_0 + e_i) > b_h] \notin [q, 1-q]$$

$$\Pr_{h\in V}[1 + w_{h,i} > 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[w_{h,i} \neq -1] \notin [q, 1-q]$$

$$\Pr_{h\in V}[w_{h,i} = -1] \notin [q, 1-q]$$

Since $\Pr_{h\in V}[w_{h,i} = 1] + \Pr_h[w_{h,i} = 0] + \Pr_h[w_{h,i} = 1] = 1$,

$$\Pr_{h\in V}[w_{h,i} = 0] \notin [q, 1-2q]$$

Thus, each index is either mostly 1, mostly 0, or mostly -1 for elements in S (since $q < 1/3$). Define $m \in \{-1, 01\}^d$ such that

$$m_i = \operatorname*{argmax}_c \Pr_{h\in V}[w_{h,i} = c]$$

□

Note that $m_1 = 1$.

6.7.5 Ratio between $m^{(-)}$ and $m^{(+)}$

Note,

$$\mathbb{E}_{h\in V}[w_h^{(+)}] = \sum_{i=1}^{d} \Pr[w_{h,i} = 1]$$

$$\leq (q)(d - m^{(+)}) + (1)m^{(+)} = qd + (1-q)m^{(+)}$$

$$m^{(+)} \geq \frac{1}{1-q}(\mathbb{E}_{h\in V}[w_h^{(+)}] - qd)$$

Further note,

$$\mathbb{E}_{h\in V}[w_h^{(+)}] = \sum_{i=1}^{d} \Pr[w_{h,i} = 1]$$

$$\geq (0)(d - m^{(+)}) + (1-q)m^{(+)} = (1-q)m^{(+)}$$

$$m^{(+)} \leq \frac{1}{1-q}\mathbb{E}_{h\in V}[w_h^{(+)}]$$

We have similar equations for $m^{(-)}$ and $\mathbb{E}_{h\in V}[w_h^{(-)}]$

Let $\bar{m}$ be the vector of m without the first component.

Recall that we have

$$\forall h \in V : w_h^{(+)} \leq r w_h^{(-)} - \frac{1}{8}d$$

$$\mathbb{E}_{h\in V}[w_h^{(+)}] \leq r\mathbb{E}_{h\in V}[w_h^{(-)}] - qrd$$

$$\frac{1}{1-q}\mathbb{E}_{h\in V}[w_h^{(+)}] \leq r\frac{1}{1-q}(\mathbb{E}_{h\in V}[w_h^{(-)}] - qd)$$

$$m^{(+)} \leq r m^{(-)}$$

$$\bar{m}^{(+)} \leq r\bar{m}^{(-)}$$

Also, recall,

$$\forall h \in V : w_h^{(-)} \leq r(w_h^{(+)} - 1) - \frac{1}{8}d$$

$$\mathbb{E}_{h\in V}[w_h^{(-)}] \leq r\mathbb{E}_{h\in V}[w_h^{(+)}] - qrd - r$$

$$\frac{1}{1-q}\mathbb{E}_{h\in V}[w_h^{(-)}] \leq r\frac{1}{1-q}(\mathbb{E}_{h\in V}[w_h^{(+)}] - qd) - \frac{r}{1-q}$$

$$m^{(-)} \leq r m^{(+)} - \frac{r}{1-q}$$

$$m^{(-)} \leq r(m^{(+)} - 1)$$

$$\bar{m}^{(-)} \leq r\bar{m}^{(+)}$$

6.7.6 Partition

Let $\bar{w}$ be the vector of w without the first component.

Definition 6.1. *Let*

- $\mathcal{X}^+ = \{x : \Pr_{h\in V}[\bar{w_h} \cdot x \geq 1] > 1 - q\}$
- $\mathcal{X}^0 = \{x : \Pr_{h\in V}[\bar{w_h} \cdot x = 0] > 1 - 2q\}$
- $\mathcal{X}^- = \{x : \Pr_{h\in V}[\bar{w_h} \cdot x \leq -1] > 1 - q\}$

Lemma 6.5. $\mathcal{X}^+, \mathcal{X}^0, \mathcal{X}^-$ *is a partition of* $\{0,1\}^{d-1}$

Proof. Since $q \leq 1/4$ and the three defining events are mutually exclusive. It is clear that A^+, A^0, A^- are disjoint. Next we show that every point is in at least one of the sets. Suppose a point x is in neither A^+ or A^-.

Using the contradiction assumption on the test $(0, x)$,

$$\Pr_{h\in V}[w_h \cdot (0,x) > 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[\bar{w_h} \cdot x > 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[\bar{w_h} \cdot x > 0] < q$$

Using the contradiction assumption on the test $(1, x)$,

$$\Pr_{h\in V}[w_h \cdot (1,x) > 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[1 + \bar{w_h} \cdot x > 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[\bar{w_h} \cdot x \geq 0] \notin [q, 1-q]$$

$$\Pr_{h\in V}[\bar{w_h} \cdot x < 0] < q$$

Combining these,

$$\Pr_{h\in V}[\bar{w_h} \cdot x = 0] = 1 - \Pr_{h\in V}[\bar{w_h} \cdot x > 0] - \Pr_{h\in V}[\bar{w_h} \cdot x < 0]$$

$$> 1 - 2q$$

Thus, $x \notin \mathcal{X}^+$ and $x \notin \mathcal{X}^-$ imply $x \in \mathcal{X}^0$ so the three sets are a partition. ☐

Definition 6.2. *Define* $\mathcal{X}^*$ *to be every* $x \in \{0,1\}^d$ *that* $(\bar{m} = 1) \cdot x = (\bar{m} = -1) \cdot x$*, where* $(\bar{m} = 1)$ *is the element-wise boolean function.*

Intuitively, this means that there are as many ones of x in positions where $\bar{m} = 1$ as there are places where $\bar{m} = -1$.

Lemma 6.6. $\mathcal{X}^* \subseteq \mathcal{X}^0$

Proof. We prove this by induction on the number of 1's in x for $x \in \mathcal{X}^*$.

The base case is $x = 0^d$ which is trivially in $\mathcal{X}^0$.

For other x, suppose $x_i = 1$ at a location where $m_i = 0$. Then we know $x - e_i \in \mathcal{X}^0$ by the induction hypothesis.

$$Pr_{h\in V}[w_h \cdot (x - e_i) = 0] > 1 - 2q$$

$$Pr_{h\in V}[w_{h,i} = 0] > 1 - 2q$$

From these,

$$Pr_{h\in V}[w_h \cdot x = 0] > 1 - 4q \geq q$$

for $q \leq 1/5$. So $x \notin \mathcal{X}^+ \cup \mathcal{X}^-$ and thus $x \in \mathcal{X}^0$.

The only other case is where $x_i = x_j = 1$ at locations where $m_i = 1$ and $m_j = -1$. Then we know $x - e_i - e_j \in \mathcal{X}^0$ from the induction hypothesis.

$$Pr_{h\in V}[w_h \cdot (x - e_i - e_j) = 0] > 1 - 2q$$

$$Pr_{h\in V}[w_{h,i} = 1] > 1 - q$$

$$Pr_{h\in V}[w_{h,j} = -1] > 1 - q$$

From these,

$$Pr_{h\in V}[w_h \cdot x = 0] > 1 - 4q \geq q$$

and similarly, $x \in \mathcal{X}^0$.

☐

6.7.7 Probability Distribution

We now define a probability distribution over $x \in \mathcal{X}^*$.

Without loss of generality, suppose $\bar{m}^{(+)} \geq \bar{m}^{(-)}$.

- Randomly draw an injection $f : \{i : \bar{m}_i = -1\} \to \{i : \bar{m}_i = 1\}$.
- Initialize $x = 0^{d-1}$
- For indices $\{i : \bar{m}_i \leq 0\}$, draw $x_i \sim$ bernoulli(1/2).
- For $\{i : \bar{m}_i = -1\}$, set $x_{f(i)} = x_i$

Note that the result $x \in \mathcal{X}^*$ because of the pairing f, there will be a 1 where $\bar{m}_i = 1$ for each 1 where $\bar{m}_i = -1$.

6.7.8 Set T

Definition 6.3. *For the probability distribution,*

$$Q(h) = \Pr_{x \in \mathcal{X}^*}[w_h \cdot x = 0]$$

Lemma 6.7. *Let $T = \{h \in V : Q(h) > 1 - 4q\}$, then $|V| > 5|T|$.*

Proof. For $x \in \mathcal{X}^*$, since $\mathcal{X}^* \subseteq \mathcal{X}^0$,

$$\Pr_{h \in V}[w_h \cdot x = 0] > 1 - 2q$$
$$\frac{\sum_{h \in V} \mathbb{1}[w_h \cdot x = 0]}{|V|} > 1 - 2q$$
$$\sum_{x \in \mathcal{X}^0} P(x) \frac{\sum_{h \in V} \mathbb{1}[w_h \cdot x = 0]}{|V|} > 1 - 2q$$
$$\frac{\sum_{h \in V} \sum_{x \in \mathcal{X}^0} P(x) \mathbb{1}[w_h \cdot x = 0]}{|V|} > 1 - 2q$$
$$\frac{\sum_{h \in V} Q(h)}{|V|} > 1 - 2q$$

$$\frac{|T|}{|V|}(1) + \frac{|V| - |T|}{|V|}(1 - 4q) > 1 - 2q$$

$$2|T| > |V|$$

□

Lemma 6.8. $|T| \leq 3$

Proof. Recall that $\bar{m}^{(+)} \leq r\bar{m}^{(-)}$ and $\bar{m}^{(-)} \leq r\bar{m}^{(+)}$ as well

Also $1 - 4q \geq 1 - \min(\frac{1}{4}, \frac{1}{2r})$ since $q \leq \min(\frac{1}{16}, \frac{1}{8r})$

For any $t \in T$, $Q(t) > 1 - 4q \geq 1 - \min(\frac{1}{4}, \frac{1}{2r})$. Define Ber(1/2) to be a Bernoulli random variable.

$$\Pr_{x \in \mathcal{X}^*}[w_t \cdot x = 0] > 1 - \min(\frac{1}{4}, \frac{1}{2r})$$
$$\mathbb{E}_f[\Pr[\sum_{i:m_i=0} w_{t,i}\text{Ber}(1/2) + ...$$
$$\sum_{i:m_i=-1} (w_{t,i} + w_{t,f(i)})\text{Ber}(1/2) = 0]] > 1 - \min(\frac{1}{4}, \frac{1}{2r})$$

Note that

$$\Pr[\sum_{i:\bar{m}_i=0} w_{t,i}\text{Ber}(1/2) + ...$$
$$\sum_{i:\bar{m}_i=-1} (w_{t,i} + w_{t,f(i)})\text{Ber}(1/2) = 0] \leq \frac{1}{2}$$

unless $\forall i : \bar{m}_i = 0 : w_{t,i} = 0$ and $\forall i : \bar{m}_i = -1 : w_{t,i} + w_{t,f(i)} = 0$, call this *condition*$(t, f)$.

$$\mathbb{E}_f[\mathbb{1}[\textit{condition}(t, f)] + ...$$
$$\frac{1}{2}(1 - \mathbb{1}[\textit{condition}(t, f)])] > 1 - \min(\frac{1}{4}, \frac{1}{2r})$$

$$\Pr_f[\textit{condition}(t, f)] > 1 - \min(\frac{1}{2}, \frac{1}{r})$$

If $\bar{m}^{(-)} = 0$, then $\bar{m}^{(+)} = 0$, and thus $\forall i : \bar{m}_i = 0 : w_{t,i} = 0$ so $t = 0^d$ and $|T| = 1 \leq 3$.

Note that $\Pr_f[\textit{condition}(t, f)] > 1/2$ implies that $\forall i : m_i = 0 : w_{t,i} = 0$.

Lemma 6.9. *If there exists i, j such that $\bar{m}_i = \bar{m}_j = -1$, then $w_{t,i} = w_{t,j}$.*

Proof. $\Pr_f[\textit{condition}(t, f)] > \frac{1}{2}$ means that

$$\Pr_f[w_{t,i} = -w_{t,f(i)}] > \frac{1}{2}$$
$$\Pr_f[w_{t,j} = -w_{t,f(j)}] > \frac{1}{2}$$

so

$$\frac{\{l : \bar{m}_l = 1 \wedge w_{t,l} = -w_{t,i}\}}{\{l : \bar{m}_l = -1\}} > 1/2$$
$$\frac{\{l : \bar{m}_l = 1 \wedge w_{t,l} = -w_{t,j}\}}{\{l : \bar{m}_l = -1\}} > 1/2$$

which is only possible if $w_{t,i} = w_{t,j}$. □

Thus, there is some $c \in \{-1, 0, 1\}$ such that $\forall i : \bar{m}_i = 1 : w_{t,i} = c$.

$$\Pr_f[\textit{condition}(t, f)] > \frac{1}{r}$$
$$\Pr_f[\forall i : \bar{m}_i = -1 : w_{t,f(i)} = -c] > 1 - \frac{1}{r}$$
$$1 - \Pr_f[\exists i : \bar{m}_i = -1 : w_{t,f(i)} \neq -c] > 1 - \frac{1}{r}$$
$$\Pr_f[\exists i : \bar{m}_i = -1 : w_{t,f(i)} \neq -c] < \frac{1}{r}$$

Suppose $\exists j : \bar{m}_j = 1 : w_{t,j} \neq -c$,

$$\Pr_f[\exists i : \bar{m}_i = -1 : f(i) = j] = \frac{1}{r}$$

which is a contradiction. So $\forall j : \bar{m}_j = 1 : w_{t,j} = -c$.

Thus, c completely determines t. Since there are three options for c, there are three options for t, and $|T| \leq 3$. $\square$

Since $|T| \leq 3$ and $2|T| \geq |V|$, $|V| \leq 6$. Thus, there is a split of 1/6 which is a contradiction since $q \leq \frac{1}{8}$. Thus, the lemma is proved. And thus the binary linear classifier problem is split-neighborly.