1. *Monotone disjunctions.*

(a) There are as many disjunctions as there are subsets of features, so $|H| = 2^d$.

(b) The true error of $h$ can be bounded thus, with probability at least $1 - \delta$:

$$\text{err}(h) \leq \frac{1}{n} \ln \frac{|H|}{\delta} = \frac{1}{n} \left( d \ln 2 + \ln \frac{1}{\delta} \right).$$

(c) $|H_k| \leq d^k$, so we get

$$\text{err}(h) \leq \frac{1}{n} \ln \frac{|H|}{\delta} = \frac{1}{n} \left( k \ln d + \ln \frac{1}{\delta} \right).$$

2. By the central limit theorem, $\hat{\nu}$ follows roughly a $N(3/4, 1/1600)$ distribution. With 95% probability, $\hat{\nu}$ will fall within 2 standard deviations of its mean, that is, in the interval $[0.7, 0.8]$.

3. *VC dimension.*

(a) $VC(H) = 2$. To prove this, we need two steps. We will first exhibit 2 points that $H$ shatters, and second prove that there is no set of 3 points that $H$ shatters.

For the first part, consider 1, 2. We can easily find intervals containing both and neither of these points. We can also find an interval $[0.5, 1.5]$ containing just 1, and $[1.5, 2.5]$ containing just 2. This covers all possibilities for labelling 1 and 2. Thus $H$ shatters $\{1, 2\}$.

Next, consider any set of 3 points. Without loss of generality, they are $a < b < c$. Observe that there is no interval that contains $a$ and $c$ but doesn’t contain $b$. Therefore, it is impossible to label $a$ 1, $c$ 1, and $b$ 0. This implies $H$ does not shatter $\{a, b, c\}$. Since $\{a, b, c\}$ was arbitrary, we are done.

(b) $VC(H) = 4$. Consider the points $(1, 0), (0, 1), (-1, 0), (0, -1)$. These can be shattered by $H$. Thus $VC(H) \geq 4$. It now suffices to show that $VC(H) < 5$. To do this, we must show that for any 5 points, $H$ does not shatter them.

Let $S$ be a set of 5 points. Let $a$ be the point with largest $x$-coordinate, $b$ be the point with largest $y$-coordinate, $c$ be the point with smallest $x$-coordinate, and $d$ be the point with smallest $y$ coordinate. Note that $a, b, c, d$ are not necessarily distinct. Let $e$ be any point in $S$ that is not $a, b, c, d$. Suppose for $H$ shatters $S$. Then there must exist $h$ such that $h(a) = h(b) = h(c) = h(d) = 1$ and $h(e) = 0$. However, this is impossible. A rectangle that contains $a, b, c, d$ has an x-coordinate range and a y-coordinate range that encompass all of $S$, and thus include $e$, which gives us a contradiction.

4. It suffices to show that for any integer $n$ that there exists a set of $n$ points that $H$ shatters. Fix $n$, and consider the set $S$ of $n$ evenly spaced points on the unit circle. We claim that $H$ shatters this set.

To prove this, consider any $T \subset S$. The key observation is that the convex hull of $T$ includes all points in $T$ and includes no points in $S \setminus T$. This is because we can connect the points of $T$ in a convex polygon, and note that this polygon only intersects the unit circle at the points of $T$. Therefore the classifier in $H$ corresponding to the convex hull of $T$ (i.e. classifying all points in this region 1 and all other points 0) classifies $T$ as 1 and $S \setminus T$ as 0. Since $T$ was arbitrary, this shows $H$ shatters $S$.

5. Let $C_1 = C_2$ both be the class of intervals described in problem 3(a). Then $\max(d_1, d_2) = 2$ since $d_1 = d_2 = 2$. However, consider 3 points $a < b < c$ on the real line. We can see that $C = C_1 \cup C_2$ shatters $\{a, b, c\}$, since we can just take 2 interval, one covering $a$ and one covering $c$ to handle the case where $a, b, c$ are labeled 1, 0, 1. All other cases can be done with a single interval (meaning we pick the same interval from $C_1$ and $C_2$).
6. (a) We observe that for any $u$, $\sigma(u) - \sigma(-u) = u$. Therefore, $a^\top x + b$ can be written as $1 \cdot \sigma(a^\top x + b) + (-1) \cdot \sigma(-a^\top x - b)$ – which corresponds to a single hidden layer neural network with $k = 2$.

(b) Notice that $\sigma(u) + \sigma(-u) = |u|$. Let $c_1 = c_2 = 1$, $a_1 = -a_2 = a$, $b_1 = -b_2 = b$, and $e = 0$. Then, $f(x) = \sigma(a^\top x + b) + \sigma(-a^\top x - b) = |a^\top x + b|$, which is not linear.

7. The set of one-hidden-layer ReLU networks is defined by

$$\mathcal{F} = \left\{ x \mapsto \sum_{i=0}^{k} c_i \sigma(a_i^\top x + b_i) + e : k \in \mathbb{N}, a_i \in \mathbb{R}^d, b_i, c_i, e \in \mathbb{R} \right\}$$

For any $f \in \mathcal{F}$, $f$ has the form

$$f(x) = \sum_{i=0}^{k} c_i \sigma(a_i^\top x + b_i) + e$$

Then,

$$g(x) = f(x) + b = \sum_{i=0}^{k} c_i \sigma(a_i^\top x + b_i) + (e + b) \Rightarrow g \in \mathcal{F}$$

$$h(x) = f(Dx) = \sum_{i=0}^{k} c_i \sigma((Da_i)^\top x + b_i) + e = \sum_{i=0}^{k} c_i \sigma((Da_i)^\top x + b_i) + e \Rightarrow h \in \mathcal{F}$$

Therefore, $\mathcal{F}$ is closed under translation and scaling.