1 Equivalence of Parameterization

There are two natural parameterizations of intervals. The first is to model the interval boundaries via:
\[
\hat{\ell} = f_t(x; w_t), \quad \hat{\ell} = f_u(x; w_u)
\]
The second is to model the interval center and size:
\[
\hat{y} = f(x; w), \quad \Delta = g(x; v)
\]
Here we show that for linear predictors, both parameterizations are equivalent. Specifically, we show that for every \((w_t, w_u)\) there exist some \((\hat{w}, \hat{v})\) whose predictions coincide on all \(x\), and vice versa.

Let \(w_t, w_u \in \mathbb{R}^d\), and consider some \(x \in \mathcal{X}\). Note that:
\[
\begin{align*}
\hat{y} &= \frac{1}{2}(w_t^\top x + w_u^\top x) = \left(\frac{1}{2}(w_t + w_u)\right)^\top x = \hat{w}^\top x \\
\Delta &= w_u^\top x - w_t^\top x = (w_u - w_t)^\top x = \hat{v}^\top x
\end{align*}
\]
Similarly, for \(w, v \in \mathbb{R}^d\), we have:
\[
\begin{align*}
\hat{\ell} &= w^\top x - \frac{1}{2}v^\top x = (w - \frac{1}{2}v)^\top x = \hat{w}_t^\top x \\
\hat{\ell} &= w^\top x + \frac{1}{2}v^\top x = (w + \frac{1}{2}v)^\top x = \hat{w}_u^\top x
\end{align*}
\]
We note that the only practical difference here would be when each component is regularized separately (e.g., with a different regularization constant).

2 VC of Functions and Losses

Recall that for a loss function \(L\) and a function class \(\mathcal{F}\), we define their composition as:
\[
L(\mathcal{F}) = \{L(y, f(x)) : f \in \mathcal{F}\}
\]
we denote the by \(L_{0/1}(\mathcal{H})\) the composition of the binary 0/1-loss function \(L_{0/1}(\cdot)\) with a binary function class \(\mathcal{H}\), where:
\[
L_{0/1}(y, \hat{y}) = 1_{\{y \neq \hat{y}\}}
\]
Here we prove the following claim:

**Lemma 1.** The VC dimension of \(\mathcal{H}\) equals the VC dimension of \(L_{0/1}(\mathcal{H})\).

**Proof.** The important observation here is that while both \(\mathcal{H}\) and \(L_{0/1}(\mathcal{H})\) include functions with binary outputs, the functions differ in their domain. Specifically, functions in \(\mathcal{H}\) map items \(x\) to binary outputs \(y \in \{0, 1\}\), while functions in \(L_{0/1}(\mathcal{H})\) take as input pairs \((x, y)\) with \(y \in \{0, 1\}\) and, via some \(h \in \mathcal{H}\), output the loss value \(z \in \{0, 1\}\).

Assume the VC dimension of \(\mathcal{H}\) is \(m\), then there exist some \(x_1, \ldots, x_m\) which shatter \(\mathcal{H}\). This means that for every \(y_1, \ldots, y_m\), there exists some \(h_y \in \mathcal{H}\) for which \(h_y(x_i) = y_i\) for every \(i\). Consider the set of pairs \((x_1, 0), \ldots, (x_m, 0)\). For any \(z_1, \ldots, z_m\), we have some \(h_z \in \mathcal{H}\) for which \(h_z(x_i) = z_i\) for every \(i\). This means that any \(z_i = 0\) gives \(h_z(x_i) = 0\), and hence:
\[
L_{0/1}(0, h_z(x_i)) = 0 = z_i
\]
Similarly, for any \(z_i = 1\) we have \(h_z(x_i) = 1\) and:
\[
L_{0/1}(0, h_z(x_i)) = 1 = z_i
\]
Now, assume the VC dimension of \(L_{0/1}(\mathcal{H})\) is \(m\). Then there exist some \((x_1, y_1), \ldots, (x_m, y_m)\) which shatter \(L_{0/1}(\mathcal{H})\). This means that for every \(z_1, \ldots, z_m\), there exists some \(h_z \in \mathcal{H}\) such that \(L_{0/1}(y_i, h_z(x_i)) = z_i\). Consider the set \(x_1, \ldots, x_m\). For any \(y_1, \ldots, y_m\), set \(z_i = 0\) for all \(i\). Hence, the corresponding \(h_z\) is such that \(L_{0/1}(y_i, h_z(x_i)) = 0\) for all \(i\), which means that \(h(x_i) = y_i\) as needed. \(\square\)