A. Supplement

A.1. Proof of technical lemmas

Proof of Lemma 1

Proof. Let $Z$ and $Z’$ be the random variables corresponding to $F(S \cup \{s\})$ and $F(S)$ respectively. Note that we have

$$F(S) = \sum_{z’ \sim Z'} \sum_{c \in \{0,1\}} \Pr[Z’ = z’, C = c] \log \frac{\Pr[Z’ = z’, C = c]}{\Pr[Z’ = z']} \Pr[C = c]$$

$$= \sum_{z’ \sim Z'} \Pr[Z’ = z’] \sum_{c \in \{0,1\}} \Pr[C = c] \Pr[Z’ = z’] \log \frac{\Pr[C = c] \Pr[Z’ = z’]}{\Pr[C = c]}$$

$$= \sum_{z’ \sim Z'} \Pr[Z’ = z’] f(\Pr[C = 0|Z’ = z’]),$$

where we have

$$f(t) = t \log \frac{t}{\Pr[C = 0]} + (1 - t) \log \frac{1 - t}{\Pr[C = 1]},$$

which is a convex function over $t \in [0, 1]$. Next, we have

$$\Delta_s F(S) = F(S \cup \{s\}) - F(S)$$

$$= \sum_{z \sim Z} \Pr[Z = z] f(\Pr[C = 0|Z = z]) - \sum_{z’ \sim Z'} \Pr[Z’ = z’] f(\Pr[C = 0|Z’ = z’])$$

$$= \Pr[Z = s'] f(\Pr[C = 0|Z = s']) + \Pr[Z = s] f(\Pr[C = 0|Z = s]) - \Pr[Z’ = s'] f(\Pr[C = 0|Z’ = s']).$$

Notice that $Z’ = s’$ implies that $Z = s$ or $Z = s’$. Hence we have $\Pr[Z’ = s’] = \Pr[Z = s'] + \Pr[Z = s]$ and

$$\Pr[C = 0|Z’ = s'] = \frac{\Pr[Z = s'] \Pr[C = 0|Z = s'] + \Pr[Z = s] \Pr[C = 0|Z = s]}{\Pr[Z = s'] + \Pr[Z = s]}.$$

Now, if we set $p = \Pr[Z = s'], q = \Pr[Z = s], \alpha = \Pr[C = 0|Z = s']$ and $\beta = \Pr[C = 0|Z = s]$, and combine the previous two inline equalities, we have

$$\Delta_s F(S) = pf(\alpha) + qf(\beta) - (p + q)f\left(\frac{p\alpha + q\beta}{p + q}\right).$$

Some Basic Tools: In Lemmas 2 and 5 we show two basic properties of convex functions that later become handy in our proof. We use the following property of convex functions to prove Lemma 2. For any convex function $f$ and any three numbers $a < b < c$ we have

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(b)}{c - b}. \quad (12)$$

Note that this also implies

$$\frac{f(c) - f(a)}{c - a} = \frac{1}{c - a} (f(c) - f(b) + f(b) - f(a))$$

$$\leq \frac{1}{c - a} (f(c) - f(b) + \frac{b - a}{c - b} (f(c) - f(b))) \quad \text{By Inequality 12}$$

$$= \frac{1}{c - a} \left(\frac{c - b + b - a}{c - b} \right)^{f(c) - f(b))}$$

$$= \frac{f(c) - f(b)}{c - b}.$$ 

(13)
Similarly we have
\[
\frac{f(c) - f(a)}{c - a} = \frac{1}{c - a} \left( f(c) - f(b) + f(b) - f(a) \right)
\geq \frac{1}{c - a} \left( \frac{c - b}{b - a} (f(b) - f(a)) + f(b) - f(a) \right)
\geq \frac{1}{c - a} \left( \frac{c - b + b - a}{b - a} (f(b) - f(a)) \right)
= \frac{f(b) - f(a)}{b - a}.
\]
(14)

**Proof of Lemma 2:**

**Proof.** First, we prove
\[
\frac{f(p\alpha + q\gamma) - f(p\alpha + q\beta)}{q\gamma - q\beta} \leq \frac{f(\gamma) - f(\beta)}{\gamma - \beta}.
\]
(15)
Recall that $\alpha \leq \beta \leq \gamma$, and $p + q = 1$. Hence we have $p\alpha + q\beta \leq p\alpha + q\gamma, \beta \leq \gamma$. We prove Inequality 15 in two cases of $p\alpha + q\gamma \leq \beta$, and $\beta < p\alpha + q\gamma$.

**Case 1.** In this case we have $p\alpha + q\beta \leq p\alpha + q\gamma \leq \beta \leq \gamma$. we have
\[
\frac{f(p\alpha + q\gamma) - f(p\alpha + q\beta)}{q\gamma - q\beta} = \frac{f(p\alpha + q\gamma) - f(p\alpha + q\beta)}{(p\alpha + q\gamma) - (p\alpha + q\beta)}
\leq \frac{f(\beta) - f(p\alpha + q\gamma)}{\beta - (p\alpha + q\gamma)}
\leq \frac{f(\gamma) - f(\beta)}{\gamma - \beta}.
\]
By Inequality 12

**Case 2.** In this case we have $p\alpha + q\beta \leq \beta \leq p\alpha + q\gamma \leq \gamma$. we have
\[
\frac{f(p\alpha + q\gamma) - f(p\alpha + q\beta)}{q\gamma - q\beta} = \frac{f(p\alpha + q\gamma) - f(p\alpha + q\beta)}{(p\alpha + q\gamma) - (p\alpha + q\beta)}
\leq \frac{f(\beta) - f(p\alpha + q\gamma)}{(p\alpha + q\gamma) - \beta}
\leq \frac{f(\gamma) - f(\beta)}{\gamma - \beta}.
\]
By Inequality 13

Next we use Inequality 15 to prove the lemma. By multiplying both sides of Inequality 15 by $q(\gamma - \beta)$ we have
\[
f(p\alpha + q\gamma) - f(p\alpha + q\beta) \leq qf(\gamma) - qf(\beta).
\]
By rearranging the terms and adding $pf(\alpha)$ to both sides we have
\[
(pf(\alpha) + qf(\beta)) - f(p\alpha + q\beta) \leq (pf(\alpha) + qf(\gamma)) - f(p\alpha + q\gamma),
\]
as desired.

**Proof of Lemma 5:**

**Proof.** We have
\[
\frac{p + q}{p + q} f\left(\frac{p\alpha + q\beta}{p + q}\right) + \frac{q' - q}{p + q} f(\beta) \geq f\left(\frac{p + q}{p + q} \frac{p\alpha + q\beta}{p + q} + \frac{q' - q}{p + q} \beta\right)
\]
By convexity
\[
= f\left(\frac{p\alpha + q\beta}{p + q} + \frac{q' - q}{p + q} \beta\right)
\]
\[
= f\left(\frac{p\alpha + q\beta}{p + q'}\right),
\]

By multiplying both sides by $p + q'$ we have
\[
(p + q)f\left(\frac{p\alpha + q\beta}{p + q}\right) + q'f(\beta) - qf(\beta) \geq (p + q')f\left(\frac{p\alpha + q'\beta}{p + q'}\right).
\]

By rearranging the terms and adding $pf(\alpha)$ to both sides we have
\[
pf(\alpha) + qf(\beta) - (p + q)f\left(\frac{p\alpha + q\beta}{p + q}\right) \leq pf(\alpha) + q'f(\beta) - (p + q')f\left(\frac{p\alpha + q'\beta}{p + q'}\right),
\]
as desired.

A.2. Empirical Evaluation Details

We implement the neural network using TensorFlow and train it using the AdamOptimizer (Abadi et al., 2016; Kingma & Ba, 2014). The following set of neural network hyperparameters are tuned by evaluating 2000 different configurations on the hold-out set as suggested by a Gaussian Process black-box optimization routine.

<table>
<thead>
<tr>
<th>hyperparameter</th>
<th>search range</th>
</tr>
</thead>
<tbody>
<tr>
<td>hidden layer size</td>
<td>[100, 1280]</td>
</tr>
<tr>
<td>num hidden layers</td>
<td>[1, 5]</td>
</tr>
<tr>
<td>learning rate</td>
<td>[1e-6, 0.01]</td>
</tr>
<tr>
<td>gradient clip norm</td>
<td>[1.0, 1000.0]</td>
</tr>
<tr>
<td>$L_2$-regularization</td>
<td>[0, 1e-4]</td>
</tr>
</tbody>
</table>