STAT205 HW4

June 7, 2022

Isaac K-M

Note that submissions were written up with respect to the 6/6/2022 version of the homework, as opposed to that posted the morning of 6/7/2022.

Problem 1

Note that x_i here is equivalent to $x^{(i)}$, i.e. marking a training datapoint.

a.)

Recall from HW4 that we have

$$\begin{aligned} y_i &= (\Phi\theta)_i \\ &= \theta_0 + \sum_{j=1}^{n-1} \theta_j (x_i - x_j)_+ \\ &= \theta_0 + \sum_{j=1}^{n-1} \theta_j (1 \cdot x_i - x_j)_+, \end{aligned}$$

and more generally

$$\begin{aligned} \hat{y}(x) &= \theta_0 + \sum_{j=1}^{n-1} \theta_j (x - x_j)_+ \\ &= \underbrace{\theta_0}_{\Theta_0} + \sum_{j=1}^{n-1} \underbrace{\theta_j}_{\Theta_{w_2,j}} (\underbrace{1}_{\Theta_{w_2,j}} \cdot x - \underbrace{x_j}_{\Theta_{b,j}})_+. \end{aligned}$$

This straightforwardly gives

- $\Theta_0 = \theta_0$

- $\Theta_{w_1,j} = 1$ for $j = 1, \dots n-1$
- $\Theta_{w_2,j} = \theta_j$ for $j = 1, \dots n-1$
- $\Theta_{b,j} = x_j$ for $j = 1, \dots n-1$

Hence, we can write

$$\Theta(\theta) = (\theta_0, \{x_j\}_{j=1}^{n-1}, \mathbf{1}_{n-1}, \{\theta_j\}_{j=1}^{n-1}).$$

As we see above, the relationship is 1-1, so we can similarly write the reverse via

$$\theta(\Theta)_j = \begin{cases} \Theta_0 & j = 0 \\ \Theta_{w_2,j} & j > 0 \end{cases}.$$

b.)

Recall in HW4 that the regularization scheme was given by

$$\lambda Q(\theta) = \lambda ||M\theta||_2^2 = \lambda \theta^T M^T M \theta.$$

As we showed in HW4, the entries of $M^T M$ correspond to the number of times in the training data that the relu term associated with each coefficient was activated, i.e.

$$M^T M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & n-1 & n-2 & n-3 & n-4 & \dots & 1 \\ 0 & n-2 & n-2 & n-3 & n-4 & \dots & 1 \\ 0 & n-3 & n-3 & n-3 & n-4 & \dots & 1 \\ 0 & n-4 & n-4 & n-4 & n-4 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

However in this setting, we only have

$$\lambda Q(\theta) = \lambda ||\theta||_2^2 = \lambda \theta^T \theta,$$

i.e. there is no "weight" matrix sandwiched between the inner product of θ with itself. In this way, each θ_j^2 term in $\theta^T \theta$ contributes evenly to the penalty (this was not the case in HW4), and the solution is a proper ridge/L2 (kernel) regression.

Now, assuming we are in the same special setting where Θ has 1-1 natural correspondence with θ, with

- $\Theta_0 = \theta_0$
- $\Theta_{w_1,j} = 1$ for $j = 1, \dots n-1$
- $\Theta_{w_2,j} = \theta_j$ for $j = 1, \dots n-1$

- $\Theta_{b,j} = x_j$ for $j = 1, \ldots n-1$,

we can leverage the natural 1-1 correspondence between Θ and θ to reconstruct a ridge penalty through for Θ. Already, under this correspondence, we have

$$f(x_i; \Theta) = \phi(x_i)^T \theta,$$

so if we set

$$Q(\Theta) = \Theta_{w_2}^T \Theta_{w_2},$$

we will have on substitution

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} (y_i - f(x_i; \Theta))^2 + \lambda Q(\Theta) &= n^{-1} \sum_{i=1}^{n} (y_i - \phi(x_i)^T \theta)^2 + \lambda \Theta_{w_2}^T \Theta_{w_2} \\ &= n^{-1} ||y - \Phi\theta||_2^2 + \lambda \Theta_{w_2}^T \Theta_{w_2} \\ &= n^{-1} ||y - \Phi\theta||_2^2 + \lambda \theta^T \theta \\ &= n^{-1} ||y - \Phi\theta||_2^2 + \lambda ||\theta||_2^2. \end{aligned}$$

In other words, this will make the optimization task identical, introducing a natural isomorphism. And as the NN (special case) can be reduced to ridge regression here, such an optimization is in fact representable as a quadratic optimization (see last line above), and a solution exists in closed form for $\lambda > 0$.

c.)

Note that we are assumed to be out of the special case here, and back to a full 1-NN.

For this problem, first recall the following lemma, as set forth in L13.28:
Lemma

$$min \sum_j x_j^2 + y_j^2 \text{ subject to } x_j y_j = z_j$$

for some $\{z_j\}_{j \in J}$ is achieved when

$$x_j = y_j = \sqrt{z_j}.$$

Hence,

$$\sum_j x_j^2 + y_j^2 = 2 \sum_j |z_j|$$

in such scenarios. As an intuitive example, you should think about a unit circle.

Proof

Now, for the proof, first examine $f(x, \Theta)$. We can show that this really is just linear with respect to one set of coefficients (instead of two), i.e.

$$
\begin{aligned}
f(x;\Theta) &= \Theta_0 + \sum_{j=1}^{d_1} \Theta_{w_2,j}(\Theta_{w_1,j} \cdot x - \Theta_{b,j})_+ \\
&= \Theta_0 + \sum_{j=1}^{d_1} \Theta_{w_2,j}\Theta_{w_1,j}(x - \Theta_{b,j}/\Theta_{w_1,j})_+ \\
&= \tilde{\theta}_0 + \sum_{j=1}^{d_1} \tilde{\theta}_j (x - \tilde{b}_j)_+ \\
&= \tilde{\phi}_{\tilde{b}}(x)^T \tilde{\theta},
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{\phi}_{\tilde{b}}(x)^T &= [1, (x-\tilde{b}_1)_+, (x-\tilde{b}_2)_+, \ldots, (x-\tilde{b}_{d_1})_+] \\
&= \left[1, \left(x - \frac{\Theta_{b,1}}{\Theta_{w_1,1}}\right)_+, \left(x - \frac{\Theta_{b,2}}{\Theta_{w_1,2}}\right)_+, \ldots, \left(x - \frac{\Theta_{b,d_1}}{\Theta_{w_1,d_1}}\right)_+\right]
\end{aligned}
$$

akin to HW4. This is the linear prediction with which we're familiar, albeit with new knots $\{\tilde{b}_j\}_{j=1}^{d_1}$ and weights $\{\tilde{\theta}_j\}_{j=0}^{d_1}$. That is, the dual-weighted Θ setup has an isomorphism with a single-weighted linear setup.

We can then vectorize the MSE portion of the objective to

$$
\begin{aligned}
\sum_{i=1}^{N} (y_i - f(x_i;\Theta))^2 &= \sum_{i=1}^{N} (y_i - \tilde{\phi}_{\tilde{b}}(x_i)^T \tilde{\theta})^2 \\
&= ||y - \Phi_\Theta \tilde{\theta}||_2^2,
\end{aligned}
$$

where (again, we'll keep with the zero-indexing of columns and coefficients) the basis is

$$
\Phi_{\Theta i,j} = \begin{cases} 1 & j = 0 \\ \left(x_i - \frac{\Theta_{b,j}}{\Theta_{w_1,j}}\right)_+ & j = 1, \ldots, d_1 \end{cases}
$$

Thus, the takeaway is that minimization of

$$
\sum_{i=1}^{N} (y_i - f(x_i;\Theta))^2
$$

is identical to the minimization of

$$
||y - \Phi_\Theta \tilde{\theta}||_2^2,
$$

where Φ_Θ has been defined in terms of Θ and the training dataset $\{x_i\}_{i=1}^n$ above.

Now, suppose we have found our Θ_{opt}, and suppose that this Θ_{opt} achieves an MSE of

$$\mathcal{M}(x, \Theta_{opt}) = \sum_{i=1}^{n} (y_i - f(x_i; \Theta_{opt}))^2.$$

By our work above, we have

$$\begin{aligned}\mathcal{M}(x, \Theta_{opt}) &= n^{-1} \sum_{i=1}^{n} (y_i - f(x_i; \Theta_{opt}))^2 \\ &= n^{-1} ||y - \Phi_{\Theta_{opt}} \tilde{\theta}_{opt}||_2^2.\end{aligned}$$

Then (knowing the answer ahead of time), we can rewrite the full minimization

$$\min_\Theta \mathcal{M}(x, \Theta) + \lambda Q(\Theta)$$

as

$$\min_\Theta \lambda(||\Theta_{w_1}||_2^2 + ||\Theta_{w_2}||_2^2) = \min_\Theta \lambda(\sum_{j=1}^{d_1} \Theta_{w_1,j}^2 + \Theta_{w_2,j}^2)$$

subject to $\mathcal{M}(x, \Theta) = \mathcal{M}(x, \Theta_{opt})$. As shown in the relu simplification above, we have

$$\tilde{\theta}_j = \Theta_{w_1,j} \Theta_{w_2,j};$$

hence, we invoke the lemma, and see that this minimization is the same as

$$\min_{\tilde{\theta}} \lambda(\sum_{j=1}^{d_1} |\tilde{\theta}_j|) = \lambda ||\tilde{\theta}||_1.$$

subject to $\mathcal{M}(x, \Theta) = \mathcal{M}(x, \Theta_{opt})$. However, we know from above that

$$\mathcal{M}(x, \Theta) = \mathcal{M}(x, \Theta_{opt}) \implies n^{-1}||y - \Phi_\Theta \tilde{\theta}||_2^2 = \mathcal{M}(x, \Theta_{opt}).$$

Thus, in all, we see that the Θ_{opt} that satisfies

$$\min_\Theta \left[\mathcal{M}(x, \Theta) + \lambda(||\Theta_{w_1}||_2^2 + ||\Theta_{w_2}||_2^2) \right]$$

corresponds directly to the $\tilde{\theta}_{opt}$ that satisfies

$$\min_{\tilde{\theta}} \left[n^{-1}||y - \Phi_{\Theta_{opt}} \tilde{\theta}||_2^2 + \lambda ||\tilde{\theta}||_1 \right],$$

where again we have required $n^{-1}||y - \Phi_{\Theta_{opt}} \tilde{\theta}||_2^2 = \mathcal{M}(x; \Theta_{opt})$. In other words, it can be reconfigured as a Lasso. And since the Lasso can be solved via quadratic programming methods, we have a quadratic optimization over linear space. Lastly, I would consider the correspondence here to be non-linear, as a critical step involves computing a $\Phi_{\Theta_{opt}}$ via repeated applications of the ReLU, which is the most well-known *non-linear* activation function. In other words, construction of the design matrix to extract $\tilde{\theta}$ requires non-linear activity, so I would consider the correspondence non-linear.

Problem 2

a.)

Simple addition, subtraction, and multiplication properties give

$$\phi_{b_2}(\phi_{b_1}(x)) = \begin{cases} relu(x - (b_1 + b_2)) & b_2 \geq 0 \\ relu(x - b_1) & b_2 < 0. \end{cases}$$

An intuitive way to think of this is that if $b_2 > 0$, it imposes more stringent standards on the ReLU; otherwise, it adds no standards, so the standards imposed by b_1 suffice.

b.)

Here, it is helpful to split into cases. First, consider what happens when x survives the first activation, i.e. $x > b_1$. We have

$$x > b_1 \implies w_1 \cdot \underbrace{relu(x - b_1)}_{>0} = w_1(x - b_1).$$

In order to satisfy the second activation, we then need

$$w_1(x - b_1) - b_2 > 0 \implies x > b_1 + \frac{b_2}{w_1}.$$

If this is satisfied, all activations are survived; hence, we have a composition of linear functions and

$$\phi_{b_3}(x) = w_2(w_1(x - b_1) - b_2).$$

and zero otherwise.

Second, consider what happens when $x < b_1$, i.e. x does not survive the first activation. This will give

$$w_1(x - b_1) = 0,$$

so the rest of the function rests solely on b_2. If $b_2 \leq 0$, then

$$w_1(x - b_1) - b_2 = 0 - b_2 > 0$$

so the second relu will activate; otherwise, when $b_2 \geq 0$ the second relu will not activate, and return zero.

At this point, we have solved the survival logic for the activations. If we combine everything in indicators, in all we have

$$\phi_{b_3}(x) = \mathbf{1}(x > b_1)\mathbf{1}(x > b_1 + b_2/w_1)[w_2(w_1(x - b_1) - b_2)] \\ + \mathbf{1}(x \leq b_1)\mathbf{1}(b_2 < 0)[-w_2 \cdot b_2].$$

c.)

Note that the result above is derived without consideration for the sign of w_1, w_2, so the expression holds.

d.)

f_ℓ **Non-decreasing**
First, consider $x' > x$. For any $b_i^{(\ell)}$, it is necessarily the case that

$$relu(x' - b_i^{(\ell)}) \geq relu(x - b_i^{(\ell)}).$$

Then, since $c_i^{(\ell)} \geq 0$, we have

$$c_i^{(\ell)} \cdot relu(x' - b_i^{(\ell)}) \geq c_i^{(\ell)} \cdot relu(x - b_i^{(\ell)})$$

and hence

$$\sum_i c_i^{(\ell)} \cdot relu(x' - b_i^{(\ell)}) \geq \sum_i c_i^{(\ell)} \cdot relu(x - b_i^{(\ell)}),$$

as desired.

f_ℓ **Increasing**
First, let $\tilde{i}$ be the i such that b_i^ℓ satisfies $\min\{b_i^\ell : c_i^\ell > 0\}$.

By construction, $b_{\tilde{i}}^\ell$ is the term (for which the coefficient is nonzero) which spends the least time at zero; i.e. it's the most "easily" activated. Hence, for any $x > x_{0,\ell}$, we have

$$relu(x - x_{0,\ell}) = relu(x - b_{\tilde{i}}^\ell) > 0 \implies c_{\tilde{i}}^\ell \cdot relu(x - x_{0,\ell}) = c_{\tilde{i}}^\ell \cdot relu(x - b_{\tilde{i}}^\ell) > 0$$

Then, for any $x' > x$, we similarly have (just as before, since $x' > x_{0,\ell}$)

$$c_{\tilde{i}}^\ell \cdot relu(x' - x_{0,\ell}) = c_{\tilde{i}}^\ell \cdot relu(x' - b_{\tilde{i}}^\ell) > 0$$

And since we have: (i) $x' > x$; (ii) $c_{\tilde{i}}^\ell \cdot relu(x - x_{0,\ell}) > 0$; and (iii) $c_{\tilde{i}}^\ell \cdot relu(x' - x_{0,\ell}) > 0$, we are guaranteed

$$c_{\tilde{i}}^\ell \cdot relu(x' - x_{0,\ell}) > c_{\tilde{i}}^\ell \cdot relu(x - x_{0,\ell}),$$

or

$$c_{\tilde{i}}^\ell \cdot relu(x' - b_{\tilde{i}}^\ell) > c_{\tilde{i}}^\ell \cdot relu(x - b_{\tilde{i}}^\ell).$$

From before, we know that for all other $i \neq \tilde{i}$

$$c_i^\ell \cdot relu(x' - b_i^\ell) > c_i^\ell \cdot relu(x - b_i^\ell).$$

This then gives

$$\begin{aligned}
\sum_{i=1}^{N} c_i^\ell \cdot relu(x' - b_i^\ell) &= c_{\tilde{i}}^\ell \cdot relu(x' - b_{\tilde{i}}^\ell) + \sum_{i \neq \tilde{i}} c_i^\ell \cdot relu(x' - b_i^\ell) \\
&> c_{\tilde{i}}^\ell \cdot relu(x - b_{\tilde{i}}^\ell) + \sum_{i \neq \tilde{i}} c_i^\ell \cdot relu(x' - b_i^\ell) \\
&\geq c_{\tilde{i}}^\ell \cdot relu(x - b_{\tilde{i}}^\ell) + \sum_{i \neq \tilde{i}} c_i^\ell \cdot relu(x - b_i^\ell) \\
&= \sum_{i=1}^{N} c_i^\ell \cdot relu(x - b_i^\ell),
\end{aligned}$$

showing the desired increase.

f_3 Non-decreasing
First, observe

$$\begin{aligned}
f_3(x) &= f_2(f_1(x)) \\
&= \sum_{i=1}^{n_2} c_i^2 \phi_{b_i^2} \left(\sum_{j=1}^{n_1} c_j^1 \phi_{b_i^1}(x) \right)
\end{aligned}$$

Now, suppose that $x' > x$, but that $f_3(x') < f_3(x)$. By the construction above, this necessarily implies that there exists some i^* such that

$$c_i^{*,2} \phi_{b_i^{*,2}} \left(\sum_{j=1}^{n_1} c_j^1 \phi_{b_i^1}(x) \right) < 0.$$

However, since $c_i^{*,2} > 0$ and since the $\phi_{b_i^{*,2}} : \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$, we necessarily have a contradiction – that is, it is impossible for this term to be sub-zero. This gives non-decreasing.

f_3 Increasing

First, if it is the case that $x_{0,3} = x_{0,1}$, then increasingness on $x > x_{0,3} = x_{0,1}$ follows from the increasingness proof above: we know that $f_1 = c_i^1 \phi_{b_i^1}(x)$ is increasing on this domain. Then, since f_2 is non-decreasing, we have over $x > x_{0,3} = x_{0,1}$ a non-decreasing function of an increasing function, which is increasing.

Second, if it is the case that $x_{0,3} = f_1^{-1}(x_{0,2}$, then we have (reintroducing our $\tilde{i}$ notation for the bias that satisfies the minimum).

$$\begin{aligned}
f_1^{-1}(x_{0,2}) &= \sup\{x : f_1(x) \leq x_{0,2}\} \\
&= \sup\{x : \sum_{i=1}^{n_1} c_1^1 \cdot relu(x - b_i^1) \leq x_{0,2}\} \\
&= \sup\{x : \sum_{i=1}^{n_1} c_1^1 \cdot relu(x - b_i^1) \leq \min\{b_i^2 : c_i^2 > 0\}\} \\
&= \sup\{x : \sum_{i=1}^{n_1} c_1^1 \cdot relu(x - b_i^1) \leq b_{\tilde{i}}^2\}.
\end{aligned}$$

The supremum will be satisfied at equality, and hence we will have

$$x_{0,3} \quad s.t. \quad \sum_{i=1}^{n_1} c_1^1 \cdot relu(x_{0,3} - b_i^1) = f_1(x_{0,3}) = b_{\tilde{i}}^2.$$

Now, take any $x' > x \geq x_{0,3}$. Then, because we are in case two and $x_{0,3} > x_{0,1}$, we have

$$f_1(x') > f_1(x) > f_1(x_{0,3}) = b_{\tilde{i}}^2.$$

For the $\tilde{i}$ term, we have

$$c_{\tilde{i}}^2 \cdot relu(f_1(x_{0,3}) - b_{\tilde{i}}^2) = 0$$

by construction, but

$$c_{\tilde{i}}^2 \cdot relu(f_1(x) - b_{\tilde{i}}^2) > 0$$

and

$$c_{\tilde{i}}^2 \cdot relu(f_1(x') - b_{\tilde{i}}^2) > 0$$

because

$$f_1(x') > f_1(x) > f_1(x_{0,3}) = b_{\tilde{i}}^2.$$

And since $f_1(x') > f_1(x)$, we get

$$c_{\tilde{i}}^2 \cdot relu(f_1(x') - b_{\tilde{i}}^2) > c_{\tilde{i}}^2 \cdot relu(f_1(x) - b_{\tilde{i}}^2).$$

Then, for all other $i \neq \tilde{i}$, we have

$$\sum_{i \neq \tilde{i}} c_i^2 \cdot relu(f_1(x') - b_i^2) \geq \sum_{i \neq \tilde{i}} c_i^2 \cdot relu(f_1(x) - b_i^2).$$

Thus, putting everything together, we get

$$
\begin{aligned}
f_2(f_1(x')) &= \sum_{i=1}^{n_2} c_i^2 \cdot relu(f_1(x') - b_i^2) \\
&= c_{\tilde{i}}^2 \cdot relu(f_1(x') - b_{\tilde{i}}^2) + \sum_{i \neq \tilde{i}} c_i^2 \cdot relu(f_1(x') - b_i^2) \\
&> c_{\tilde{i}}^2 \cdot relu(f_1(x) - b_{\tilde{i}}^2) + \sum_{i \neq \tilde{i}} c_i^2 \cdot relu(f_1(x') - b_i^2) \\
&\geq c_{\tilde{i}}^2 \cdot relu(f_1(x) - b_{\tilde{i}}^2) + \sum_{i \neq \tilde{i}} c_i^2 \cdot relu(f_1(x) - b_i^2) \\
&= \sum_{i=1}^{n_2} c_i^2 \cdot relu(f_1(x') - b_i^2) \\
&= f_2(f_1(x)),
\end{aligned}
$$

as desired. Again, we have contained ourself to $x > x_{0,3}$ in order to introduce keep ourselves on the active area of the ReLU, in order for this to work.

f_3 **Piece-wise Linear**
At this point in the class, the following two facts should be clear:

1.) The composition of piece-wise linear function f (N_f unique nontrivial knots) with a piece-wise linear function g (N_g unique nontrivial knots) will itself be a piece-wise linear function, with as many as $N_g + N_f$ knots. If knots align perfectly, then this composition will have $N_{fg} = N_g = N_f$ knots (i.e. the same amount); if there are no shared knots across f, g, then the new composition will have $N_{fg} = N_g + N_f$. So under composition, you get a new piece-wise linear function, with as few as $\max\{N_g, N_f\}$ new knots and as many as $N_g + N_f$ new knots. Note that some of these new knots may be trivial (i.e. slope approaching from both sides is the same), depending on the composition going on on the LHS and RHS of the knot in question.

To see the mechanics of this consider the following, for $f \circ g$, where f and g are piece-wise linear with knot sets b_f, b_g. First, concatenate and sort (ascending) $\tilde{b} = sort(\{b_f, b_g\})$ to get a "new" knot set. Second, examine each $\tilde{b}_i, \tilde{b}_{i+1}$ (a "chunk"). Within this chunk, for $x \in [\tilde{b}_i, \tilde{b}_{i+1}]$, $f \circ g$ on x will be linear, since a composition of linear functions is linear. Do this over every "chunk", and each chunk will be linear; hence, a piece-wise linear function in all.

2.) The addition of piece-wise linear functions is also a piece-wise linear function. The logic/intuition is identical to that above – however now, within each chunk, we add the linear functions instead of compose them (i.e. multiply).

These two facts give us everything we need to iteratively show piece-wise linearity. Already, we know that

$$c_i^1 \cdot relu(x - b_i^1)$$

is piece-wise linear (last lecture; basic intuition). By the additive property, it follows that

$$f_1(x) \sum_i c_i^1 \cdot relu(x - b_i^1)$$

must also be piece-wise linear. Similarly, we know that

$$c_j^2 \cdot relu(z - b_j^2)$$

is piece-wise linear in z. Since a composition of piece-wise linear functions piece-wise linear, we know that

$$c_j^2 \cdot relu(f_1(x) - b_j^2)$$

is also piece-wise linear too. Lastly, using the additive property again, we have that

$$\sum_j c_j^2 \cdot relu(f_1(x) - b_j^2) = f_2(f_1(x))$$

must also be piece-wise linear, completing the proof.

Where f_3 Has Knots

Since each $relu(x - b_i^\ell)$ is a piecewise function with a knot at b_i^ℓ, the sum

$$f_1(x) = \sum_{i=1}^{n_1} c_i^1 \cdot relu(x - b_1^\ell)$$

has knots/kinks at $\{b_i^1\}$. Intuitively, any time as x crosses over b_i^1 (i.e. goes from $x < b_i^1$ to $x > b_i^1$), the function

$$\sum_{k \neq i} c_k^1 \cdot relu(x - b_k^1)$$

may be smooth in x, but the non-linear activation will still exist for the $c_i^1 \cdot relu(x - b_1^1)$, so overall the function will still have a knot there. This happens over all $i = 1, \dots, n_1$ biases; hence a guaranteed $\{b_i^1\}_{i=1}^{n_1}$ knots. No matter how else you compose f_1 with some other f_2, these initial knots – as they directly interact with the x on entry to the model – will remain. Intuitively, they'll be in on the ground floor.

Next, you may also see knots at $\{f_1^{-1}(b_j^2)$. If we expand out this term a bit, the intuition becomes clear

$$\{f_1^{-1}(b_j^2) = \sup\{x : f_1(x) \leq b_j^2\},$$

and we know that for $x : f_1(x) > b_j^2$, we will have activation

$$c_j^2 \cdot (f_1(x) - b_j^2) \neq 0,$$

whereas for $x : f_1(x) \leq b_j^2$, we will have no activation

$$c_j^2 \cdot (f_1(x) - b_j^2) = 0,$$

In other words, x crosses some threshold that, through f_1 switches on an activation two layers ahead. In this way, it introduces a piece-wise linearity (via the relu) in $f_1(x)$ at b_j^1; hence, if we take that threshold with respect to x instead of $f_1(x)$, the threshold is $f_1^{-1}(b_j^1)$.

Of course, not all of these points will come to pass as additional knots. For instance, $f_1(x)$ maybe configured such that activation of

$$relu(f_1(x) - b_2^j)$$

is impossible, in which case no knot would be added. So it is possible, but not certain, to add knots in this manner.

With the knot sets for f_2 and f_1 established above, it is clear that the knot set for their composition, f_3 is

$$\{b_i^1\} \cup \{f_1^{-1}(b_j^2)\}.$$

The first set $\{b_i^1\}$ follows from the first commentary above; the second set $\{f_1^{-1}(b_j^2)\}$ follows from the second half, where x induces a threshold crossing/activation in the second layer, and hence a knot. In the event that such a threshold crossing is induced in all n_2 functions in the second layer, this will tack on an additional n_2 knots, giving $n_1 + n_2$ knots in all.

Conclude/Explain

Our results above incrementally and logically lead us to the stated conclusion. The first part of the sentence is given by the shown piecewise linearity above; the second part of the sentence is given by the N_g, N_f discussion in the piecewise linearity proof. That is, when you compose or add a linear piecewise function with N_f knots with a linear piecewise function with N_g knots, there are $N_f + N_g$ "chunks" to evaluate the composition over, occurring at the combined set of all knots. This leaves ample room for $N_f + N_g$ total knots; however, it is likely that the combined piece-wise linearities – either through the relu, multiplication, or addition, that some of these knots will be trivialized (e.g. LHS slope = RHS slope), so there may well be fewer knots.

e.)

See Colab pdf.

3.)

See Colab pdf.

STAT205_HW5_supp

June 7, 2022

[1]:
```
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install pypandoc
```

```
Reading package lists… Done
Building dependency tree
Reading state information… Done
pandoc is already the newest version (1.19.2.4~dfsg-1build4).
pandoc set to manually installed.
The following package was automatically installed and is no longer required:
  libnvidia-common-460
Use 'apt autoremove' to remove it.
The following additional packages will be installed:
  fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre
  javascript-common libcupsfilters1 libcupsimage2 libgs9 libgs9-common
  libijs-0.35 libjbig2dec0 libjs-jquery libkpathsea6 libpotrace0 libptexenc1
  libruby2.5 libsynctex1 libtexlua52 libtexluajit2 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-did-you-mean ruby-minitest
  ruby-net-telnet ruby-power-assert ruby-test-unit ruby2.5
  rubygems-integration t1utils tex-common tex-gyre texlive-base
  texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-recommended texlive-pictures texlive-plain-generic tipa
Suggested packages:
  fonts-noto apache2 | lighttpd | httpd poppler-utils ghostscript
  fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic
  | fonts-ipafont-gothic fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri
  ruby-dev bundler debhelper gv | postscript-viewer perl-tk xpdf-reader
  | pdf-viewer texlive-fonts-recommended-doc texlive-latex-base-doc
  python-pygments icc-profiles libfile-which-perl
  libspreadsheet-parseexcel-perl texlive-latex-extra-doc
  texlive-latex-recommended-doc texlive-pstricks dot2tex prerex ruby-tcltk
  | libtcltk-ruby texlive-pictures-doc vprerex
The following NEW packages will be installed:
  fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre
  javascript-common libcupsfilters1 libcupsimage2 libgs9 libgs9-common
  libijs-0.35 libjbig2dec0 libjs-jquery libkpathsea6 libpotrace0 libptexenc1
  libruby2.5 libsynctex1 libtexlua52 libtexluajit2 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-did-you-mean ruby-minitest
  ruby-net-telnet ruby-power-assert ruby-test-unit ruby2.5
```

```
  rubygems-integration t1utils tex-common tex-gyre texlive texlive-base
  texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-extra texlive-latex-recommended texlive-pictures
  texlive-plain-generic texlive-xetex tipa
0 upgraded, 47 newly installed, 0 to remove and 45 not upgraded.
Need to get 146 MB of archives.
After this operation, 460 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-droid-fallback
all 1:6.0.1r16-1.1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-lato all 2.0-2
[2,698 kB]
Get:3 http://archive.ubuntu.com/ubuntu bionic/main amd64 poppler-data all
0.4.8-2 [1,479 kB]
Get:4 http://archive.ubuntu.com/ubuntu bionic/main amd64 tex-common all 6.09
[33.0 kB]
Get:5 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-lmodern all
2.004.5-3 [4,551 kB]
Get:6 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-noto-mono all
20171026-2 [75.5 kB]
Get:7 http://archive.ubuntu.com/ubuntu bionic/universe amd64 fonts-texgyre all
20160520-1 [8,761 kB]
Get:8 http://archive.ubuntu.com/ubuntu bionic/main amd64 javascript-common all
11 [6,066 B]
Get:9 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsfilters1
amd64 1.20.2-0ubuntu3.1 [108 kB]
Get:10 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsimage2
amd64 2.2.7-1ubuntu2.9 [18.6 kB]
Get:11 http://archive.ubuntu.com/ubuntu bionic/main amd64 libijs-0.35 amd64
0.35-13 [15.5 kB]
Get:12 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjbig2dec0 amd64
0.13-6 [55.9 kB]
Get:13 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9-common
all 9.26~dfsg+0-0ubuntu0.18.04.16 [5,093 kB]
Get:14 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9 amd64
9.26~dfsg+0-0ubuntu0.18.04.16 [2,265 kB]
Get:15 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjs-jquery all
3.2.1-1 [152 kB]
Get:16 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libkpathsea6
amd64 2017.20170613.44572-8ubuntu0.1 [54.9 kB]
Get:17 http://archive.ubuntu.com/ubuntu bionic/main amd64 libpotrace0 amd64
1.14-2 [17.4 kB]
Get:18 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libptexenc1
amd64 2017.20170613.44572-8ubuntu0.1 [34.5 kB]
Get:19 http://archive.ubuntu.com/ubuntu bionic/main amd64 rubygems-integration
all 1.11 [4,994 B]
Get:20 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 ruby2.5 amd64
2.5.1-1ubuntu1.11 [48.6 kB]
Get:21 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby amd64 1:2.5.1
```

```
[5,712 B]
Get:22 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 rake all
12.3.1-1ubuntu0.1 [44.9 kB]
Get:23 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-did-you-mean all
1.2.0-2 [9,700 B]
Get:24 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-minitest all
5.10.3-1 [38.6 kB]
Get:25 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:26 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-power-assert all
0.3.0-1 [7,952 B]
Get:27 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-test-unit all
3.2.5-1 [61.1 kB]
Get:28 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libruby2.5
amd64 2.5.1-1ubuntu1.11 [3,072 kB]
Get:29 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libsynctex1
amd64 2017.20170613.44572-8ubuntu0.1 [41.4 kB]
Get:30 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtexlua52
amd64 2017.20170613.44572-8ubuntu0.1 [91.2 kB]
Get:31 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtexluajit2
amd64 2017.20170613.44572-8ubuntu0.1 [230 kB]
Get:32 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libzzip-0-13
amd64 0.13.62-3.1ubuntu0.18.04.1 [26.0 kB]
Get:33 http://archive.ubuntu.com/ubuntu bionic/main amd64 lmodern all 2.004.5-3
[9,631 kB]
Get:34 http://archive.ubuntu.com/ubuntu bionic/main amd64 preview-latex-style
all 11.91-1ubuntu1 [185 kB]
Get:35 http://archive.ubuntu.com/ubuntu bionic/main amd64 t1utils amd64 1.41-2
[56.0 kB]
Get:36 http://archive.ubuntu.com/ubuntu bionic/universe amd64 tex-gyre all
20160520-1 [4,998 kB]
Get:37 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 texlive-
binaries amd64 2017.20170613.44572-8ubuntu0.1 [8,179 kB]
Get:38 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-base all
2017.20180305-1 [18.7 MB]
Get:39 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-fonts-
recommended all 2017.20180305-1 [5,262 kB]
Get:40 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-latex-base all
2017.20180305-1 [951 kB]
Get:41 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-latex-
recommended all 2017.20180305-1 [14.9 MB]
Get:42 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive all
2017.20180305-1 [14.4 kB]
Get:43 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-pictures
all 2017.20180305-1 [4,026 kB]
Get:44 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-latex-
extra all 2017.20180305-2 [10.6 MB]
Get:45 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-plain-
```

```
generic all 2017.20180305-2 [23.6 MB]
Get:46 http://archive.ubuntu.com/ubuntu bionic/universe amd64 tipa all 2:1.3-20
[2,978 kB]
Get:47 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-xetex all
2017.20180305-1 [10.7 MB]
Fetched 146 MB in 4s (33.9 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages …
Selecting previously unselected package fonts-droid-fallback.
(Reading database … 155632 files and directories currently installed.)
Preparing to unpack …/00-fonts-droid-fallback_1%3a6.0.1r16-1.1_all.deb …
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1) …
Selecting previously unselected package fonts-lato.
Preparing to unpack …/01-fonts-lato_2.0-2_all.deb …
Unpacking fonts-lato (2.0-2) …
Selecting previously unselected package poppler-data.
Preparing to unpack …/02-poppler-data_0.4.8-2_all.deb …
Unpacking poppler-data (0.4.8-2) …
Selecting previously unselected package tex-common.
Preparing to unpack …/03-tex-common_6.09_all.deb …
Unpacking tex-common (6.09) …
Selecting previously unselected package fonts-lmodern.
Preparing to unpack …/04-fonts-lmodern_2.004.5-3_all.deb …
Unpacking fonts-lmodern (2.004.5-3) …
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack …/05-fonts-noto-mono_20171026-2_all.deb …
Unpacking fonts-noto-mono (20171026-2) …
Selecting previously unselected package fonts-texgyre.
Preparing to unpack …/06-fonts-texgyre_20160520-1_all.deb …
Unpacking fonts-texgyre (20160520-1) …
Selecting previously unselected package javascript-common.
Preparing to unpack …/07-javascript-common_11_all.deb …
Unpacking javascript-common (11) …
Selecting previously unselected package libcupsfilters1:amd64.
Preparing to unpack …/08-libcupsfilters1_1.20.2-0ubuntu3.1_amd64.deb …
Unpacking libcupsfilters1:amd64 (1.20.2-0ubuntu3.1) …
Selecting previously unselected package libcupsimage2:amd64.
Preparing to unpack …/09-libcupsimage2_2.2.7-1ubuntu2.9_amd64.deb …
Unpacking libcupsimage2:amd64 (2.2.7-1ubuntu2.9) …
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack …/10-libijs-0.35_0.35-13_amd64.deb …
Unpacking libijs-0.35:amd64 (0.35-13) …
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack …/11-libjbig2dec0_0.13-6_amd64.deb …
Unpacking libjbig2dec0:amd64 (0.13-6) …
Selecting previously unselected package libgs9-common.
Preparing to unpack …/12-libgs9-common_9.26~dfsg+0-0ubuntu0.18.04.16_all.deb
…
```

```
Unpacking libgs9-common (9.26~dfsg+0-0ubuntu0.18.04.16) …
Selecting previously unselected package libgs9:amd64.
Preparing to unpack …/13-libgs9_9.26~dfsg+0-0ubuntu0.18.04.16_amd64.deb …
Unpacking libgs9:amd64 (9.26~dfsg+0-0ubuntu0.18.04.16) …
Selecting previously unselected package libjs-jquery.
Preparing to unpack …/14-libjs-jquery_3.2.1-1_all.deb …
Unpacking libjs-jquery (3.2.1-1) …
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack …/15-libkpathsea6_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libkpathsea6:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libpotrace0.
Preparing to unpack …/16-libpotrace0_1.14-2_amd64.deb …
Unpacking libpotrace0 (1.14-2) …
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack …/17-libptexenc1_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libptexenc1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package rubygems-integration.
Preparing to unpack …/18-rubygems-integration_1.11_all.deb …
Unpacking rubygems-integration (1.11) …
Selecting previously unselected package ruby2.5.
Preparing to unpack …/19-ruby2.5_2.5.1-1ubuntu1.11_amd64.deb …
Unpacking ruby2.5 (2.5.1-1ubuntu1.11) …
Selecting previously unselected package ruby.
Preparing to unpack …/20-ruby_1%3a2.5.1_amd64.deb …
Unpacking ruby (1:2.5.1) …
Selecting previously unselected package rake.
Preparing to unpack …/21-rake_12.3.1-1ubuntu0.1_all.deb …
Unpacking rake (12.3.1-1ubuntu0.1) …
Selecting previously unselected package ruby-did-you-mean.
Preparing to unpack …/22-ruby-did-you-mean_1.2.0-2_all.deb …
Unpacking ruby-did-you-mean (1.2.0-2) …
Selecting previously unselected package ruby-minitest.
Preparing to unpack …/23-ruby-minitest_5.10.3-1_all.deb …
Unpacking ruby-minitest (5.10.3-1) …
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack …/24-ruby-net-telnet_0.1.1-2_all.deb …
Unpacking ruby-net-telnet (0.1.1-2) …
Selecting previously unselected package ruby-power-assert.
Preparing to unpack …/25-ruby-power-assert_0.3.0-1_all.deb …
Unpacking ruby-power-assert (0.3.0-1) …
Selecting previously unselected package ruby-test-unit.
Preparing to unpack …/26-ruby-test-unit_3.2.5-1_all.deb …
Unpacking ruby-test-unit (3.2.5-1) …
Selecting previously unselected package libruby2.5:amd64.
Preparing to unpack …/27-libruby2.5_2.5.1-1ubuntu1.11_amd64.deb …
Unpacking libruby2.5:amd64 (2.5.1-1ubuntu1.11) …
```

```
Selecting previously unselected package libsynctex1:amd64.
Preparing to unpack …/28-libsynctex1_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libsynctex1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libtexlua52:amd64.
Preparing to unpack …/29-libtexlua52_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libtexlua52:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
…/30-libtexluajit2_2017.20170613.44572-8ubuntu0.1_amd64.deb …
Unpacking libtexluajit2:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libzzip-0-13:amd64.
Preparing to unpack …/31-libzzip-0-13_0.13.62-3.1ubuntu0.18.04.1_amd64.deb …
Unpacking libzzip-0-13:amd64 (0.13.62-3.1ubuntu0.18.04.1) …
Selecting previously unselected package lmodern.
Preparing to unpack …/32-lmodern_2.004.5-3_all.deb …
Unpacking lmodern (2.004.5-3) …
Selecting previously unselected package preview-latex-style.
Preparing to unpack …/33-preview-latex-style_11.91-1ubuntu1_all.deb …
Unpacking preview-latex-style (11.91-1ubuntu1) …
Selecting previously unselected package t1utils.
Preparing to unpack …/34-t1utils_1.41-2_amd64.deb …
Unpacking t1utils (1.41-2) …
Selecting previously unselected package tex-gyre.
Preparing to unpack …/35-tex-gyre_20160520-1_all.deb …
Unpacking tex-gyre (20160520-1) …
Selecting previously unselected package texlive-binaries.
Preparing to unpack …/36-texlive-
binaries_2017.20170613.44572-8ubuntu0.1_amd64.deb …
Unpacking texlive-binaries (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package texlive-base.
Preparing to unpack …/37-texlive-base_2017.20180305-1_all.deb …
Unpacking texlive-base (2017.20180305-1) …
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack …/38-texlive-fonts-recommended_2017.20180305-1_all.deb …
Unpacking texlive-fonts-recommended (2017.20180305-1) …
Selecting previously unselected package texlive-latex-base.
Preparing to unpack …/39-texlive-latex-base_2017.20180305-1_all.deb …
Unpacking texlive-latex-base (2017.20180305-1) …
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack …/40-texlive-latex-recommended_2017.20180305-1_all.deb …
Unpacking texlive-latex-recommended (2017.20180305-1) …
Selecting previously unselected package texlive.
Preparing to unpack …/41-texlive_2017.20180305-1_all.deb …
Unpacking texlive (2017.20180305-1) …
Selecting previously unselected package texlive-pictures.
Preparing to unpack …/42-texlive-pictures_2017.20180305-1_all.deb …
```

```
Unpacking texlive-pictures (2017.20180305-1) …
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack …/43-texlive-latex-extra_2017.20180305-2_all.deb …
Unpacking texlive-latex-extra (2017.20180305-2) …
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack …/44-texlive-plain-generic_2017.20180305-2_all.deb …
Unpacking texlive-plain-generic (2017.20180305-2) …
Selecting previously unselected package tipa.
Preparing to unpack …/45-tipa_2%3a1.3-20_all.deb …
Unpacking tipa (2:1.3-20) …
Selecting previously unselected package texlive-xetex.
Preparing to unpack …/46-texlive-xetex_2017.20180305-1_all.deb …
Unpacking texlive-xetex (2017.20180305-1) …
Setting up libgs9-common (9.26~dfsg+0-0ubuntu0.18.04.16) …
Setting up libkpathsea6:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up libjs-jquery (3.2.1-1) …
Setting up libtexlua52:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up fonts-droid-fallback (1:6.0.1r16-1.1) …
Setting up libsynctex1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up libptexenc1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up tex-common (6.09) …
update-language: texlive-base not installed and configured, doing nothing!
Setting up poppler-data (0.4.8-2) …
Setting up tex-gyre (20160520-1) …
Setting up preview-latex-style (11.91-1ubuntu1) …
Setting up fonts-texgyre (20160520-1) …
Setting up fonts-noto-mono (20171026-2) …
Setting up fonts-lato (2.0-2) …
Setting up libcupsfilters1:amd64 (1.20.2-0ubuntu3.1) …
Setting up libcupsimage2:amd64 (2.2.7-1ubuntu2.9) …
Setting up libjbig2dec0:amd64 (0.13-6) …
Setting up ruby-did-you-mean (1.2.0-2) …
Setting up t1utils (1.41-2) …
Setting up ruby-net-telnet (0.1.1-2) …
Setting up libijs-0.35:amd64 (0.35-13) …
Setting up rubygems-integration (1.11) …
Setting up libpotrace0 (1.14-2) …
Setting up javascript-common (11) …
Setting up ruby-minitest (5.10.3-1) …
Setting up libzzip-0-13:amd64 (0.13.62-3.1ubuntu0.18.04.1) …
Setting up libgs9:amd64 (9.26~dfsg+0-0ubuntu0.18.04.16) …
Setting up libtexluajit2:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up fonts-lmodern (2.004.5-3) …
Setting up ruby-power-assert (0.3.0-1) …
Setting up texlive-binaries (2017.20170613.44572-8ubuntu0.1) …
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
```

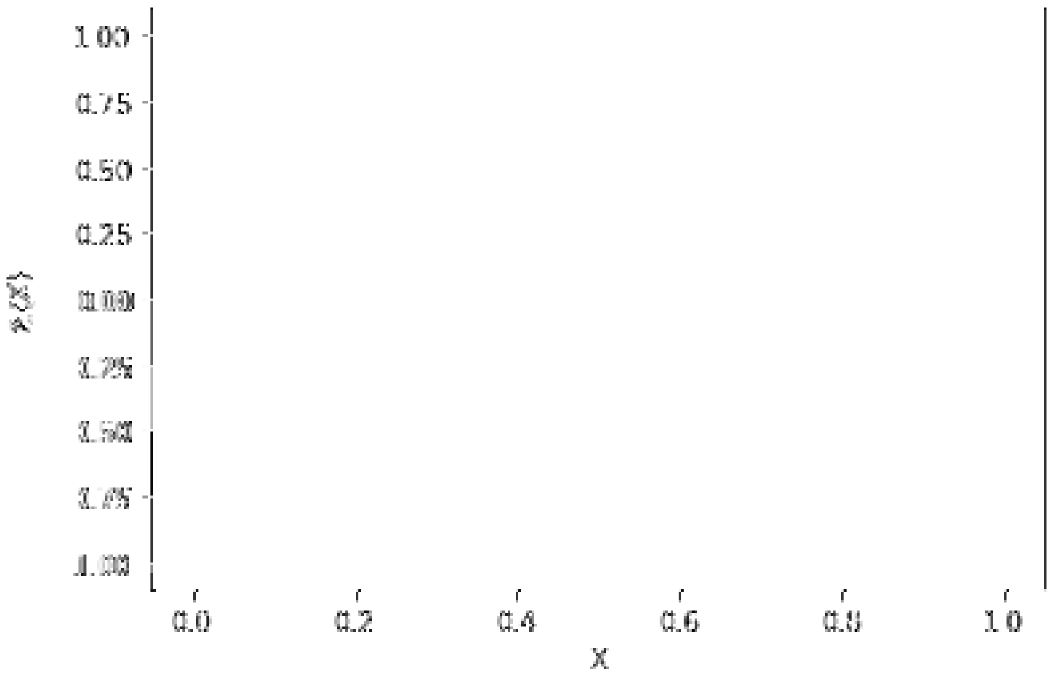
```
(bibtex) in auto mode
Setting up texlive-base (2017.20180305-1) …
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST…
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN…
mktexlsr: Updating /var/lib/texmf/ls-R…
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4:
/var/lib/texmf/tex/generic/config/pdftexconfig.tex
Setting up texlive-fonts-recommended (2017.20180305-1) …
Setting up texlive-plain-generic (2017.20180305-2) …
Setting up texlive-latex-base (2017.20180305-1) …
Setting up lmodern (2.004.5-3) …
Setting up texlive-latex-recommended (2017.20180305-1) …
Setting up texlive-pictures (2017.20180305-1) …
Setting up tipa (2:1.3-20) …
Regenerating '/var/lib/texmf/fmtutil.cnf-DEBIAN'… done.
Regenerating '/var/lib/texmf/fmtutil.cnf-TEXLIVEDIST'… done.
update-fmtutil has updated the following file(s):
        /var/lib/texmf/fmtutil.cnf-DEBIAN
        /var/lib/texmf/fmtutil.cnf-TEXLIVEDIST
If you want to activate the changes in the above file(s),
you should run fmtutil-sys or fmtutil.
Setting up texlive (2017.20180305-1) …
Setting up texlive-latex-extra (2017.20180305-2) …
Setting up texlive-xetex (2017.20180305-1) …
Setting up ruby2.5 (2.5.1-1ubuntu1.11) …
Setting up ruby (1:2.5.1) …
Setting up ruby-test-unit (3.2.5-1) …
Setting up rake (12.3.1-1ubuntu0.1) …
Setting up libruby2.5:amd64 (2.5.1-1ubuntu1.11) …
Processing triggers for mime-support (3.60ubuntu1) …
Processing triggers for libc-bin (2.27-3ubuntu1.3) …
/sbin/ldconfig.real: /usr/local/lib/python3.7/dist-
packages/ideep4py/lib/libmkldnn.so.0 is not a symbolic link

Processing triggers for man-db (2.8.3-2ubuntu0.1) …
Processing triggers for fontconfig (2.12.6-0ubuntu2) …
Processing triggers for tex-common (6.09) …
Running updmap-sys. This may take some time… done.
Running mktexlsr /var/lib/texmf … done.
Building format(s) --all.
        This may take some time… done.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
```

```
wheels/public/simple/
Collecting pypandoc
  Downloading pypandoc-1.8.1-py3-none-any.whl (20 kB)
Installing collected packages: pypandoc
Successfully installed pypandoc-1.8.1
```

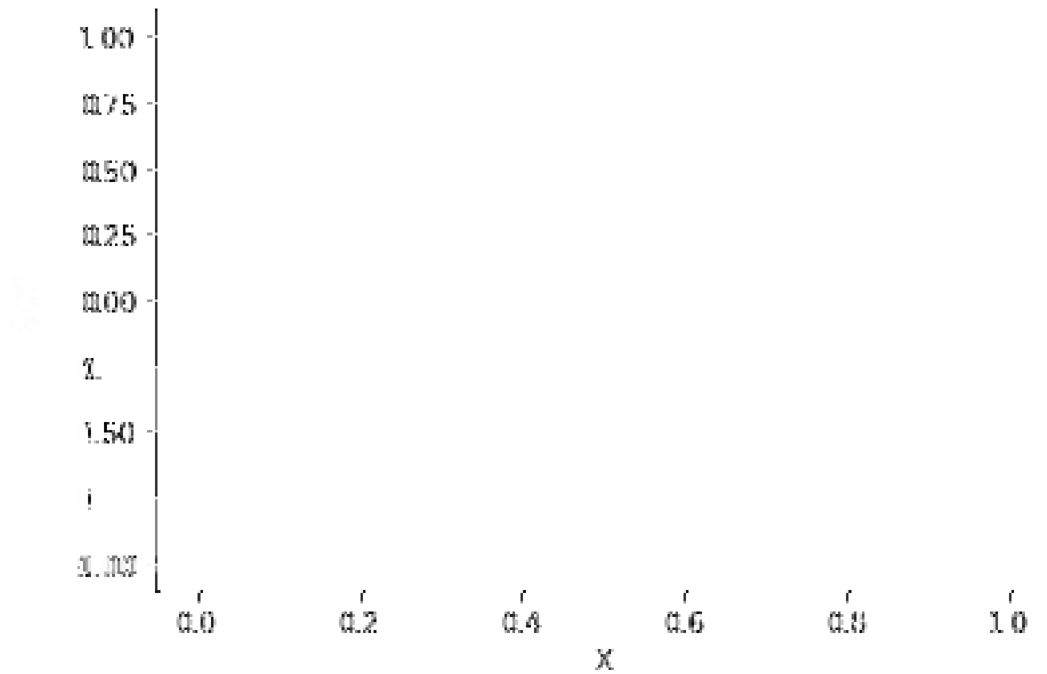
[2]:
```python
import numpy as np
import matplotlib.pyplot as plt
```

0.0.1 2E

[3]:
```python
### make toy dataset, with interpolating zeros halfway between
N = 25
X = np.arange(0, (2 * N) + 1) / (2 * N)
Y = -np.zeros_like(X)
Y[0::4] = -1.
Y[2::4] = 1.

plt.plot(X, Y)
plt.scatter(X, Y)
plt.xlabel("X")
plt.ylabel("$f_1(X)$")
plt.show()
```



[4]:
```
plt.plot(X, 2 * np.abs(Y) - 1)
plt.scatter(X, 2 * np.abs(Y) - 1)
plt.xlabel("X")
plt.ylabel("$f_3(X)$")
plt.show()
```



Piecewise Linearity: As we see, f_3 remains piecewise linear -- we've simply folded the linear interpolation that was f_1 (also piecewise linear) on itself.

Knot Count: The number of knots increases from $N+1$ (for the original $0, \ldots, N$) to $(N+1)+N$, with the additional zeros (originally part of the interpolation, but become knots on folding) added via f_2.

Additional Commentary: Here, f_3 has not only let the x_i retain their positions as knots in the linear interpolation; it's also introduced knot's at the original $f_1(x) = 0$ points, as the L1 norm makes these zeros "turnaround" points for the otherwise function -- i.e. (prior to scaling down) the L1 norm guarantees that all other transformed points, descend to, before turning around and beginning their ascent. In other words, the $2|x|-1$ function makes all zeros knots, which effectively doubles the number of knots in the interpolation.

Explanation

Sentence 1: "*A Two-Hidden Layer ReLu Deepnet on x R with any coefficients whatsoever can still be represented as a piecewise linear spline.*" This follows from identical argumentation to that outlined in the f_3 piecewise linear question above: if you just keep adding and composing piecewise linear functions, you'll get a piecewise linear function at every step. Then, over the whole dataset, you'll have a "spline."

Sentence 2: "*If some coefficients of f1 or f2 are nonpositive, the spline might involve more knots than would have been required in the positive-coefficient case*" This follows from the initial parts of problem 2 -- when you flip the sign, you may squash part of the output to zero, or you may unsquash part of the output from zero. This act of squashing/unsquashing will induce a kink, and hence another potential knot.

Sentence 3: "*Letting n_i denote the number of knots in f_i , it might involve roughly as many as (n_1 + 1) · (n_2 + 1) knots*". First, in f_1, we'll have approximately n_1 knots from the first layer: we get this by summing n_1 piecewise linear functions/relus. Then, for each $j = 1, \dots, n_2$, we examine

$$c_j^2 \cdot relu(f_1(x) - b_j^2).$$

This will take the n_1 knots from f_1, and potentially adjust each not (depending on c_j^2, b_j^2) to get n_1 "updated" knots that are specific to the j index. Do this n_2 times for each j, and now you have up to n_2 sets of n_1 "j-specific" knots. Hence, it's $\mathcal{O}(n_1 \cdot n_2)$ total knots.

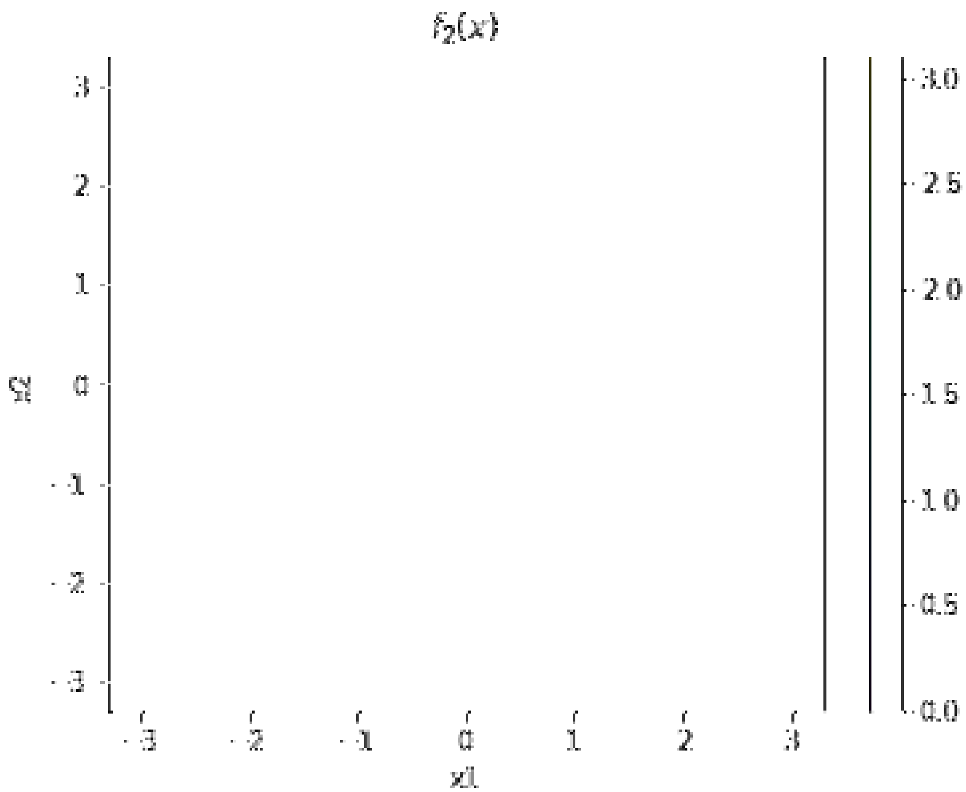
3.)

A.)

[5]:
```
import itertools
import math

plt.rcParams["figure.figsize"] = (7, 7)

U_fn = lambda i: np.array([
    math.cos(2 * math.pi * (i - 1) / 3),
    math.sin(2 * math.pi * (i - 1) / 3)
])

U = np.vstack([U_fn(1), U_fn(2), U_fn(3)])

X = np.vstack(
    [
     (i, j)
     for (i, j)
     in itertools.product(
         np.linspace(-3, 3, 250),
         np.linspace(-3, 3, 250)
      )
    ]
)

def relu(x):
  x = x.copy()
  x[x < 0] = 0.
  return x

def f1(x):
```

```
  x = x.copy()
  x[:, 0] -= 1
  x[:, 1] += 1
  return relu(x)

def f2(x):
  return relu(x @ (U.T) - 1.).sum(axis=1)
```

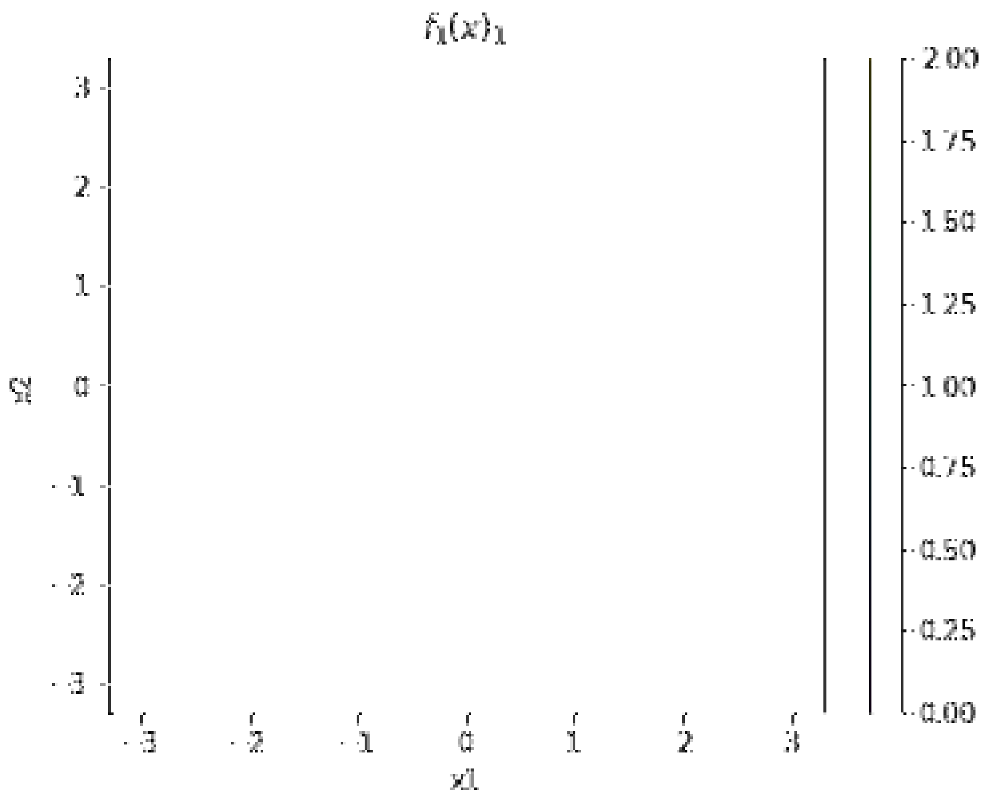
[6]:
```
plt.rcParams["figure.figsize"] = (6, 4.5)

y2 = f2(X)

plt.scatter(X[:, 0], X[:, 1], c=y2)
plt.colorbar()
plt.xlabel("x1")
plt.ylabel("x2")
plt.title("$f_2(x)$")
plt.show()
```

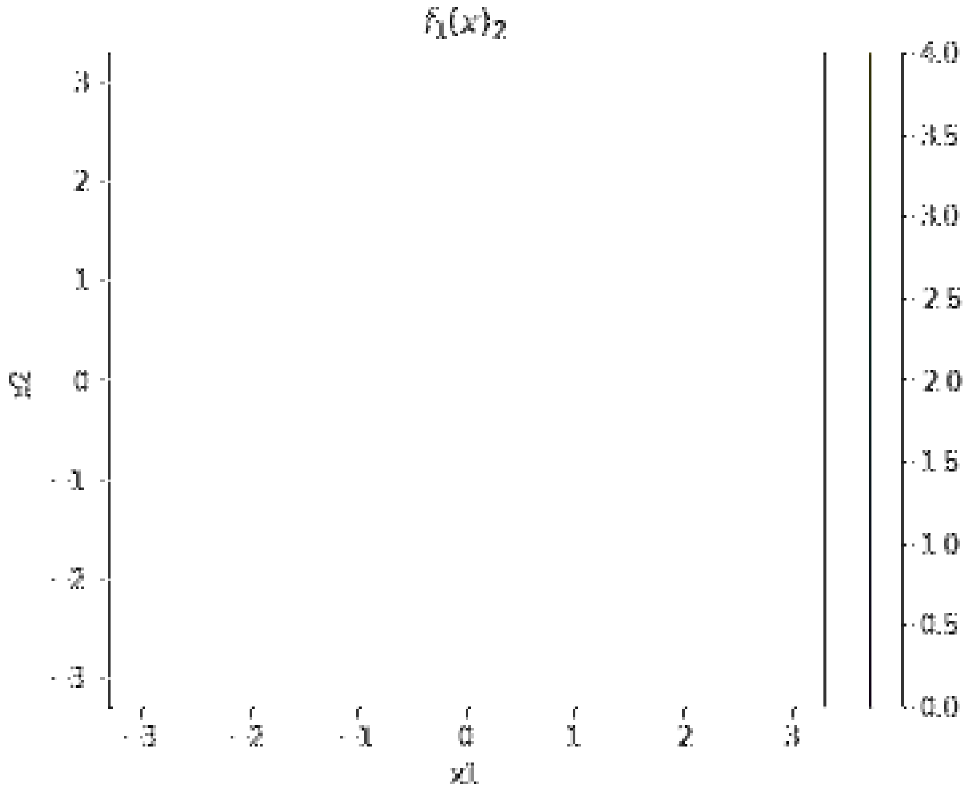


0.0.2 B.)

[7]:

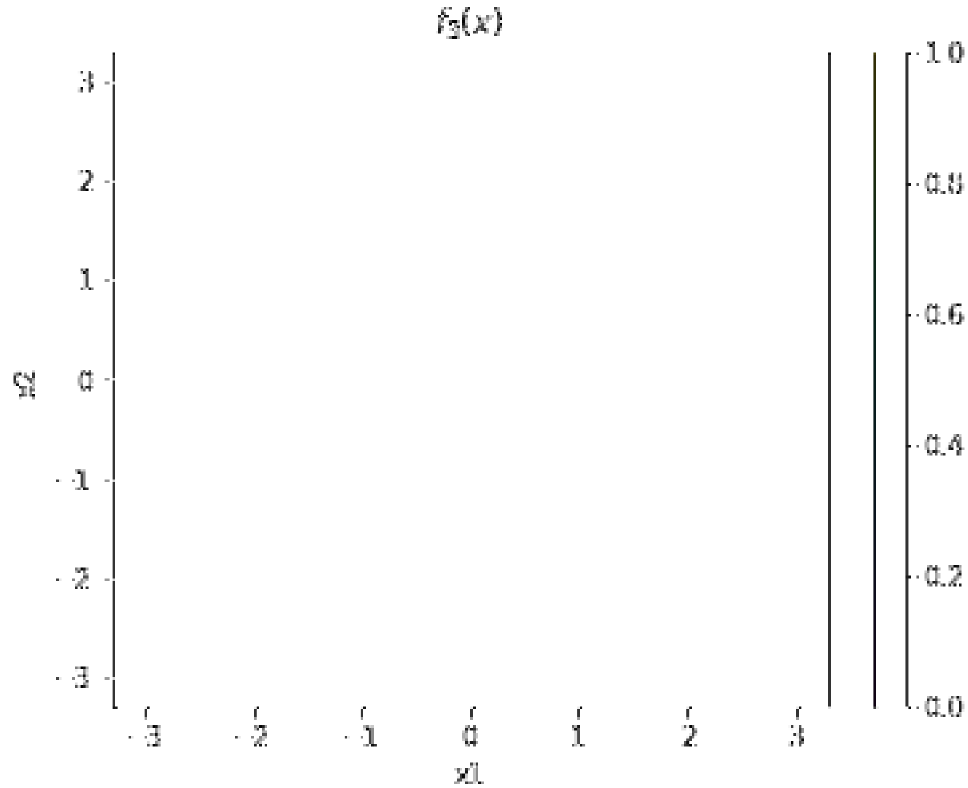
```python
y1 = f1(X)

plt.scatter(X[:, 0], X[:, 1], c=y1[:, 0])
plt.colorbar()
plt.xlabel("x1")
plt.ylabel("x2")
plt.title("$f_1(x)_1$")
plt.show()

plt.scatter(X[:, 0], X[:, 1], c=y1[:, 1])
plt.colorbar()
plt.xlabel("x1")
plt.ylabel("x2")
plt.title("$f_1(x)_2$")
plt.show()
```



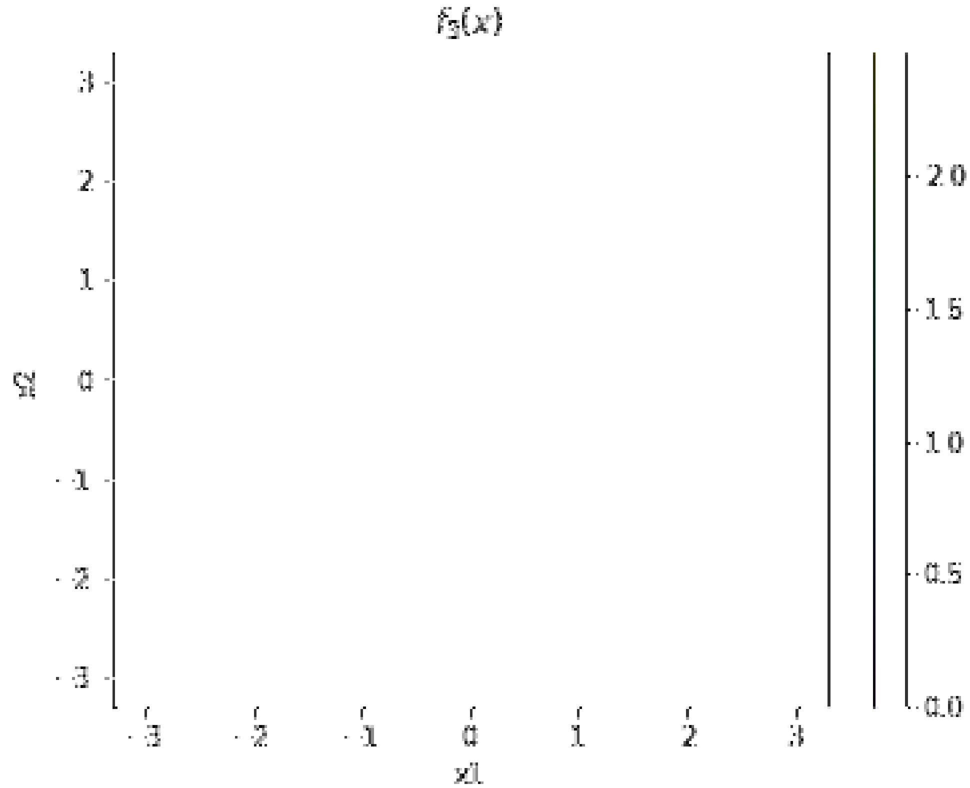


0.0.3 3C.

[8]:
```
y3 = (f2(f1(X)) > 0 ).astype(int)

plt.scatter(X[:, 0], X[:, 1], c=y3)
plt.colorbar()
plt.xlabel("x1")
plt.ylabel("x2")
plt.title("$f_3(x)$")
plt.show()


y3 = f2(f1(X))

plt.scatter(X[:, 0], X[:, 1], c=y3)
plt.colorbar()
plt.xlabel("x1")
plt.ylabel("x2")
plt.title("$f_3(x)$")
plt.show()
```



$f_3(x)$
x2
x1



0.0.4 D.)

Given the simplicity of the ReLU, the cell decomposition is straightforward:

- The line $x_2 = -1$ splits $f_1(x)_2$ into two cells.
- The line $x_1 = 1$ splits $f_1(x)_1$ into two cells.

0.0.5 E.)

Observe

$$
\begin{align}
f_2(x) &= Relu(cos(0)x_1 + sin(0)x_2 - 1) \tag{1}\\
&\quad + Relu(cos(2\pi/3)x_1 + sin(2\pi/3)x_2 - 1) \tag{2}\\
&\quad + Relu(cos(4\pi/3)x_1 + sin(4\pi/3)x_2 - 1) \tag{3}\\
&= Relu(x_1 - 1) \tag{4}\\
&\quad + Relu(-\frac{1}{2}x_1 + \frac{\sqrt{3}}{2}x_2 - 1) \tag{5}\\
&\quad + Relu(-\frac{1}{2}x_1 - \frac{\sqrt{3}}{2}x_2 - 1). \tag{6}
\end{align}
$$

From this, we see that the cell boundaries (i.e. activation thresholds) must correspond to the lines

- $x_1 - 1 > 0 \implies x_1 = 1$
- $-\frac{1}{2}x_1 + \frac{\sqrt{3}}{2}x_2 - 1 > 0 \implies x_2 = \frac{x+2}{\sqrt{3}}$
- $-\frac{1}{2}x_1 - \frac{\sqrt{3}}{2}x_2 - 1 > 0 \implies x_2 = -\frac{x+2}{\sqrt{3}}$.

Hence the lines

$$\left\{x_1 = -1, x_2 = \frac{x+2}{\sqrt{3}}, x_2 = -\frac{x+2}{\sqrt{3}}\right\}$$

define the cell decomposition.

0.0.6 F.)

As we are free to choose any subset/region A on $\mathbb{R}^2$, consider $A = \{x : x_1 < 1, x_2 < -1\}$. By the properties of f_1, we shall always have for $x \in A$

$$f_1(x) = [Relu(x_1 - 1), Relu(x_2 + 1)] = [0, 0].$$

In turn, we are guaranteed

$$f_2([0, 0]) = Relu(0 - 1) + Relu(0 - 1) + Relu(0 - 1) = 0$$

, so for any $x \in A = \{x : x_1 < 1, x_2 < -1\}$, we will have

$$f_3(0) = 0 + \mathbf{0}^T x.$$

[ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

[ ]:
```
!cp "drive/MyDrive/Colab Notebooks/STAT205_HW5_supp.ipynb" ./
```

[ ]:
```
!jupyter nbconvert --to PDF "STAT205_HW5_supp.ipynb"
```