

Neural networks with backprop library

Justin Le

The *backprop* library performs backpropagation over a *heterogeneous* system of relationships. It does so by letting you build an explicit graph and keeps track of what nodes depend on what. Let's use it to build neural networks!

Repository source is on [github](https://github.com/mstksg/backprop)¹, and so are the rendered unstable docs².

```
{-# LANGUAGE DeriveGeneric      #-}
{-# LANGUAGE GADTs              #-}
{-# LANGUAGE LambdaCase        #-}
{-# LANGUAGE RankNTypes        #-}
{-# LANGUAGE ScopedTypeVariables #-}
{-# LANGUAGE StandaloneDeriving #-}
{-# LANGUAGE TypeApplications   #-}
{-# LANGUAGE TypeInType        #-}
{-# LANGUAGE TypeOperators      #-}
{-# LANGUAGE ViewPatterns       #-}
{-# OPTIONS_GHC -fno-warn-orphans #-}

import      Data.Functor
import      Data.Kind
import      Data.Maybe
import      Data.Singletons
import      Data.Singletons.Prelude
import      Data.Singletons.TypeLits
import      Data.Type.Combinator
import      Data.Type.Product
import      GHC.Generics              (Generic)
import      Numeric.Backprop
import      Numeric.Backprop.Iso
import      Numeric.Backprop.Op
import      Numeric.LinearAlgebra.Static hiding (dot)
import      System.Random.MWC
import qualified Generics.SOP          as SOP
```

Ops

First, we define values of `Op` for the operations we want to do. `Ops` are bundles of functions packaged with their heterogeneous gradients. For simple numeric functions, *backprop* can derive `Ops` automatically. But for matrix operations, we have to derive them ourselves.

¹<https://github.com/mstksg/backprop>

²<https://mstksg.github.io/backprop>

The types help us with matching up the dimensions, but we still need to be careful that our gradients are calculated correctly.

`L` and `R` are matrix and vector types from the great *hmatrix* library.

First, matrix-vector multiplication:

```
matVec
  :: (KnownNat m, KnownNat n)
  => Op '[ L m n, R n ] (R m)
matVec = op2' $ \m v -> ( m #> v
                          , \ (fromMaybe 1 -> g) ->
                            (g `outer` v, tr m #> g)
                          )
```

Now, dot products:

```
dot :: KnownNat n
    => Op '[ R n, R n ] Double
dot = op2' $ \x y -> ( x <.> y
                      , \case Nothing -> (y, x)
                          Just g -> (konst g * y, x * konst g)
                      )
```

And for kicks, we can show an auto-derived logistic function op:

```
logistic :: Floating a => Op '[a] a
logistic = op1 $ \x -> 1 / (1 + exp (-x))
```

That's really it!

A Simple Complete Example

At this point, we already have enough to train a simple single-hidden-layer neural network:

```
simpleOp
  :: (KnownNat m, KnownNat n, KnownNat o)
  => R m
  -> BPOp s '[ L n m, R n, L o n, R o ] (R o)
simpleOp inp = withInps $ \(w1 :< b1 :< w2 :< b2 :< Ø) -> do
  -- First layer
  y1 <- matVec -$ (w1 :< x1 :< Ø)
  z1 <- op2 (+) -$ (y1 :< b1 :< Ø)
  x2 <- logistic -$ only z1
  -- Second layer
  y2 <- matVec -$ (w2 :< x2 :< Ø)
  z2 <- op2 (+) -$ (y2 :< b2 :< Ø)
  logistic -$ only z2
  where
    x1 = constRef inp
```

Now `simpleOp` can be “run” with the input vectors and parameters (`a L n m`, `R n`, `L o n`, and `R o`) and calculate the output of the neural net.

```
runSimple
  :: (KnownNat m, KnownNat n, KnownNat o)
  => R m
```

```

-> Tuple '[ L n m, R n, L o n, R o ]
-> R o
runSimple inp = runBPOp (simpleOp inp)

```

But, in defining `simpleOp`, we also generated a graph that *backprop* can use to do backpropagation, too!

```

simpleGrad
  :: forall m n o. (KnownNat m, KnownNat n, KnownNat o)
  => R m
  -> R o
  -> Tuple '[ L n m, R n, L o n, R o ]
  -> Tuple '[ L n m, R n, L o n, R o ]
simpleGrad inp targ params = gradBPOp opError params
  where
    opError :: BPOp s '[ L n m, R n, L o n, R o ] Double
    opError = do
      res <- simpleOp inp
      err <- op2 (-) -$ (res :< t      :< ∅)
      dot          -$ (err :< err :< ∅)
    where
      t = constRef targ

```

The result is the gradient of the input tuple’s components, with respect to the `Double` result of `opError` (the squared error). We can then use this gradient to do gradient descent.

With Parameter Containers

This method doesn’t quite scale, because we might want to make networks with multiple layers and parameterize networks by layers. Let’s make some basic container data types to help us organize our types, including a recursive `Network` type that lets us chain multiple layers.

```

data Layer :: Nat -> Nat -> Type where
  Layer :: { _lWeights :: L m n
            , _lBiases  :: R m
            }
            -> Layer n m
  deriving (Show, Generic)

data Network :: Nat -> [Nat] -> Nat -> Type where
  N∅    :: !(Layer a b) -> Network a '[ ] b
  (:&)  :: !(Layer a b) -> Network b bs c -> Network a (b ': bs) c

```

A `Layer n m` is a layer taking an `n`-vector and returning an `m`-vector. A `Network a '[b, c, d] e` would be a `Network` that takes in an `a`-vector and outputs an `e`-vector, with hidden layers of sizes `b`, `c`, and `d`.

Isomorphisms

The *backprop* library lets you apply operations on “parts” of data types (like on the weights and biases of a `Layer`) by using `Iso`’s (isomorphisms), like the ones from the *lens* library. The library doesn’t depend on *lens*, but it can use the `Isos` from the library and also custom-defined ones.

First, we can auto-generate isomorphisms using the *generics-sop* library:

```
instance SOP.Generic (Layer n m)
```

And then can create isomorphisms by hand for the two Network constructors:

```
netExternal :: Iso' (Network a '[] b) (Tuple '[Layer a b])
netExternal = iso (\case NØ x      -> x ::< Ø)
                  (\case I x :< Ø -> NØ x    )

netInternal :: Iso' (Network a (b '[: bs] c) (Tuple '[Layer a b, Network b bs c])
netInternal = iso (\case x :& xs      -> x ::< xs ::< Ø)
                  (\case I x :< I xs :< Ø -> x :& xs    )
```

An `Iso' a (Tuple as)` means that an `a` can really just be seen as a tuple of `as`.

Running a network

Now, we can write the `BPOp` that represents running the network and getting a result. We pass in a `Sing bs` (a singleton list of the hidden layer sizes) so that we can “pattern match” on the list and handle the different network constructors differently.

```
netOp
  :: forall s a bs c. (KnownNat a, KnownNat c)
  => Sing bs
  -> BPOp s '[ R a, Network a bs c ] (R c)
netOp sbs = go sbs
  where
    go :: forall d es. KnownNat d
        => Sing es
        -> BPOp s '[ R d, Network d es c ] (R c)
    go = \case
      SNil -> withInps $ \(x :< n :< Ø) -> do
        -- peek into the NØ using netExternal iso
        l :< Ø <- netExternal #<~ n
        -- run the 'layerOp' op, with x and l as inputs
        layerOp ~$ (x :< l :< Ø)
      SNat `SCons` ses -> withInps $ \(x :< n :< Ø) -> withSingI ses $ do
        -- peek into the (:&) using the netInternal iso
        l :< n' :< Ø <- netInternal #<~ n
        -- run the 'layerOp' BP, with x and l as inputs
        z <- layerOp ~$ (x :< l :< Ø)
        -- run the 'go ses' BP, with z and n as inputs
        go ses ~$ (z :< n' :< Ø)
    layerOp
      :: forall d e. (KnownNat d, KnownNat e)
      => BPOp s '[ R d, Layer d e ] (R e)
    layerOp = withInps $ \(x :< l :< Ø) -> do
      -- peek into the layer using the gTuple iso, auto-generated with SOP.Generic
      w :< b :< Ø <- gTuple #<~ l
      y <- matVec -$ (w :< x :< Ø)
      z <- op2 (+) -$ (y :< b :< Ø)
      logistic -$ only z
```

There’s some singletons work going on here, but it’s fairly standard singletons stuff. From *backprop* specifically, `(#<~)` lets you “split” an input ref with the given iso, and `(~$)` lets you “run” an BP within an BP, by

plugging in its inputs.

Gradient Descent

Now we can do simple gradient descent. Defining an error function:

```
err
  :: KnownNat m
  => R m
  -> BPrep s rs (R m)
  -> BPOp s rs Double
err targ r = do
  d <- opRef2 r t $ op2 (-)
  opRef2 d d      $ dot
  where
    t = constRef targ
```

And now, we can use `backprop` to generate the gradient, and shift the `Network`! Things are made a bit cleaner from the fact that `Network a bs c` has a `Num` instance, so we can use `(-)` and `(*)` etc.

```
train
  :: (KnownNat a, SingI bs, KnownNat c)
  => Double
  -> R a
  -> R c
  -> Network a bs c
  -> Network a bs c
train r x t n = case backprop (err t =<< netOp sing) (x ::< n ::< 0) of
  (_, _ :< I g :< 0) -> n - (realToFrac r * g)
```

((::<) is cons and 0 is nil for tuples.)

Main

`main`, which will train on sample data sets, is still in progress! Right now it just generates a random network using the *mwc-random* library and prints each internal layer.

```
main :: IO ()
main = withSystemRandom $ \g -> do
  n <- uniform @ (Network 4 '[3,2] 1) g
  void $ traverseNetwork sing (\l -> l <$ print l) n
```

Appendix: Boilerplate

And now for some typeclass instances and boilerplates unrelated to the *backprop* library that makes our custom types easier to use.

```
instance KnownNat n => Variate (R n) where
  uniform g = randomVector <$> uniform g <*> pure Uniform
  uniformR (l, h) g = (\x -> x * (h - l) + l) <$> uniform g

instance (KnownNat m, KnownNat n) => Variate (L m n) where
```

```

uniform g = uniformSample <$> uniform g <*> pure 0 <*> pure 1
uniformR (l, h) g = (\x -> x * (h - l) + l) <$> uniform g

instance (KnownNat n, KnownNat m) => Variate (Layer n m) where
uniform g = subtract 1 . (* 2) <$> (Layer <$> uniform g <*> uniform g)
uniformR (l, h) g = (\x -> x * (h - l) + l) <$> uniform g

instance (KnownNat m, KnownNat n) => Num (Layer n m) where
Layer w1 b1 + Layer w2 b2 = Layer (w1 + w2) (b1 + b2)
Layer w1 b1 - Layer w2 b2 = Layer (w1 - w2) (b1 - b2)
Layer w1 b1 * Layer w2 b2 = Layer (w1 * w2) (b1 * b2)
abs (Layer w b) = Layer (abs w) (abs b)
signum (Layer w b) = Layer (signum w) (signum b)
negate (Layer w b) = Layer (negate w) (negate b)
fromInteger x = Layer (fromInteger x) (fromInteger x)

instance (KnownNat m, KnownNat n) => Fractional (Layer n m) where
Layer w1 b1 / Layer w2 b2 = Layer (w1 / w2) (b1 / b2)
recip (Layer w b) = Layer (recip w) (recip b)
fromRational x = Layer (fromRational x) (fromRational x)

instance (KnownNat a, SingI bs, KnownNat c) => Variate (Network a bs c) where
uniform g = genNet sing (uniform g)
uniformR (l, h) g = (\x -> x * (h - l) + l) <$> uniform g

genNet
  :: forall f a bs c. (Applicative f, KnownNat a, KnownNat c)
  => Sing bs
  -> (forall d e. (KnownNat d, KnownNat e) => f (Layer d e))
  -> f (Network a bs c)
genNet sbs f = go sbs
  where
    go :: forall d es. KnownNat d => Sing es -> f (Network d es c)
    go = \case
      SNil          -> N0 <$> f
      SNat `SCons` ses -> (:&) <$> f <*> go ses

mapNetwork0
  :: forall a bs c. (KnownNat a, KnownNat c)
  => Sing bs
  -> (forall d e. (KnownNat d, KnownNat e) => Layer d e)
  -> Network a bs c
mapNetwork0 sbs f = getI $ genNet sbs (I f)

traverseNetwork
  :: forall a bs c f. (KnownNat a, KnownNat c, Applicative f)
  => Sing bs
  -> (forall d e. (KnownNat d, KnownNat e) => Layer d e -> f (Layer d e))
  -> Network a bs c
  -> f (Network a bs c)
traverseNetwork sbs f = go sbs
  where
    go :: forall d es. KnownNat d => Sing es -> Network d es c -> f (Network d es c)

```

```

go = \case
  SNil -> \case
    N0 x -> N0 <$> f x
  SNat `SCons` ses -> \case
    x :& xs -> (:&) <$> f x <*> go ses xs

mapNetwork1
  :: forall a bs c. (KnownNat a, KnownNat c)
  => Sing bs
  -> (forall d e. (KnownNat d, KnownNat e) => Layer d e -> Layer d e)
  -> Network a bs c
  -> Network a bs c
mapNetwork1 sbs f = getI . traverseNetwork sbs (I . f)

mapNetwork2
  :: forall a bs c. (KnownNat a, KnownNat c)
  => Sing bs
  -> (forall d e. (KnownNat d, KnownNat e) => Layer d e -> Layer d e -> Layer d e)
  -> Network a bs c
  -> Network a bs c
  -> Network a bs c
mapNetwork2 sbs f = go sbs
  where
    go :: forall d es. KnownNat d => Sing es -> Network d es c -> Network d es c -> Network d es c
    go = \case
      SNil -> \case
        N0 x -> \case
          N0 y -> N0 (f x y)
      SNat `SCons` ses -> \case
        x :& xs -> \case
          y :& ys -> f x y :& go ses xs ys

instance (KnownNat a, SingI bs, KnownNat c) => Num (Network a bs c) where
  (+)          = mapNetwork2 sing (+)
  (-)          = mapNetwork2 sing (-)
  (*)          = mapNetwork2 sing (*)
  negate       = mapNetwork1 sing negate
  abs          = mapNetwork1 sing abs
  signum       = mapNetwork1 sing signum
  fromInteger x = mapNetwork0 sing (fromInteger x)

instance (KnownNat a, SingI bs, KnownNat c) => Fractional (Network a bs c) where
  (/)          = mapNetwork2 sing (/)
  recip        = mapNetwork1 sing recip
  fromRational x = mapNetwork0 sing (fromRational x)

```