TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

The Educational Framework (EDF)

The Educational Framework (EDF)

The educational frameword (EDF) is 150 lines of Python-NumPy that implement a deep learning framework.

In EDF we write

$$
\begin{aligned}
y &= F(p, x) \\
z &= G(q, y, x) \\
u &= H(z) \\
\mathcal{L} &= u
\end{aligned}
$$

This is Python code where variables are bound to objects.

Kinds of Nodes

There are three kinds of nodes — inputs, parameters and computed nodes (CompNodes).

Inputs and computed nodes have a batch index. Parameters do not have a batch index.

The value of inputs and parameters are provided. The value of CompNodes is computed by the forward pass.

The EDF Framework

$$
\begin{aligned}
y &= F(p, x) \\
z &= G(q, y, x) \\
u &= H(z) \\
\mathcal{L} &= u
\end{aligned}
$$

x is an object in the class `Input`.

p and p are objects in subclasses of `Parameter`.

y is an object in the class F; z is an object in the class G; and u and $\mathcal{L}$ are the same object in the class H.

The classes F, G, and H are subclasses of `CompNode`.

The Core of EDF

```
def Forward():
    for c in CompNodes: c.forward()

def Backward(loss):
    for c in CompNodes + Parameters: c.grad = 0
    loss.grad = 1.
    for c in CompNodes[::-1]: c.backward()

def SGD():
    for p in Parameters:
        p.value -= eta*p.grad
```

$$y = F(p, x)$$

```
class F(CompNode):

    def __init__(self,p,x):
        CompNodes.append(self)
        self.x = x
        self.p = p

    def forward(self):
        self.value = ... compute the value ...

    def backward(self):
        self.x.addgrad(... compute the gradient ...)
        self.p.addgrad(... compute the gradient ...)
```

The Classes Input and CompNode

```
class Input:
    def __init__(self):
        pass
    def addgrad(self, delta):
        pass


class CompNode: #initialization is handled by the subclass
    def addgrad(self, delta):
        self.grad += delta
```

The Class Parameter

```
class Parameter:

    def __init__(self,value):
        Parameters.append(self)
        self.value = value

    def addgrad(self, delta):
        #sums over the minibatch
        self.grad += np.sum(delta, axis = 0)/nBatch
```

MLP in EDF

The following Python code constructs the computation graph of a multi-layer perceptron (NLP) with one hidden layer.

```
L1 = Sigmoid(Affine(Phi1,x))
Q = Softmax(Sigmoid(Affine(Phi2,L1))
ell = LogLoss(Q,y)
```

Here `x` and `y` are input computation nodes whose value have been set. Here `Phi1` and `Phi2` are "parameter packages" (a matrix and a bias vector in this case). We have computation node classes `Affine`, `Relu`, `Sigmoid`, `LogLoss` each of which has a forward and a backward method.

The `Sigmoid` Class

$$
\begin{aligned}
y[b,i] &= \sigma(x[b,i]) \\
y &= \frac{1}{1+e^{-x}} \\
\frac{dy}{dx} &= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= y(1-y)
\end{aligned}
$$

$$x.\mathrm{grad}[b,i] \mathrel{\texttt{+=}} y.\mathrm{grad}[b,i]y.\mathrm{value}[b,i](1-y.\mathrm{value}[b,i])$$

The `Sigmoid` Class

```
class Sigmoid:
  def __init__(self,x):
    CompNodes.append(self)
    self.x = x

  def forward(self):
    self.value = 1. / (1. + np.exp(-self.x.value))

  def backward(self):
    self.x.addgrad(self.grad*self.value*(1.-self.value))
```

The `Affine` Class

$$\tilde{y}[b,j] = \sum_i W[i,j]\, x[b,i] = xW$$

$$y[b,j] = \tilde{y}[b,j] - B[j] = \tilde{y} - B \text{ (broadcasting)}$$

$$\tilde{y}.\mathrm{grad}[b,j] \mathrel{+}= y.\mathrm{grad}[b,j]$$

$$B.\mathrm{grad}[j] \mathrel{-}= \frac{1}{B}\sum_b y.\mathrm{grad}[b,j]$$

$$x.\mathrm{grad}[b,i] \mathrel{+}= \sum_j \tilde{y}.\mathrm{grad}[b,j] W[i,j] = y.\mathrm{grad}\, W^\top$$

$$W.\mathrm{grad}[i,j] \mathrel{+}= \frac{1}{B}\sum_b \tilde{y}.\mathrm{grad}[b,j] x[b,i] = ???$$

```
class Affine(CompNode):

    def __init__(self,Phi,x):
        CompNodes.append(self)
        self.x = x
        self.Phi = Phi

    def forward(self):
        self.value = (np.matmul(self.x.value,
                                self.Phi.w.value)
                      - self.Phi.b.value)
```

```
def backward(self):

    self.x.addgrad(
       np.matmul(self.grad,
                 self.Phi.w.value.transpose()))

    self.Phi.b.addgrad(- self.grad)

    self.Phi.w.addgrad(self.x.value[:,:,np.newaxis]
                       * self.grad[:,np.newaxis,:])
```

Procedures in EDF

```
def MLP(Phi,x)

   if len(Phi) = 0
      return x

   return Sigmoid(Affine(Phi[0],MLP(Phi[1:],x)))
```

END