EE 526X Deep Machine Learning: Theory and Practice

Homework 2

Kerui Tan

Problem 1

$$
\begin{aligned}
\frac{\partial \mathbf{p}}{\partial \mathbf{z}} &= \left[\frac{\partial p_j}{\partial z_i}\right]_{i,j=1,1}^{n,n} \\
&= \left[\frac{\partial \frac{e^{z_j}}{\sum_{k=1}^{n} e^{z_k}}}{\partial z_i}\right]_{i,j=1,1}^{n,n} \\
&= \left[i = j : \frac{e^{z_j}\sum_{k=1}^{n} e^{z_k} - e^{z_i}e^{z_j}}{(\sum_{k=1}^{n} e^{z_k})^2}, i \neq j : \frac{0 - e^{z_i}e^{z_j}}{(\sum_{k=1}^{n} e^{z_k})^2}\right]_{i,j=1,1}^{n,n} \\
&= \left[i = j : \frac{e^{z_j}}{\sum_{k=1}^{n} e^{z_k}}\frac{\sum_{k=1}^{n} e^{z_k} - e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}, i \neq j : -\frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}\frac{e^{z_j}}{\sum_{k=1}^{n} e^{z_k}}\right]_{i,j=1,1}^{n,n} \\
&= \left[i = j : p_j(\frac{\sum_{k=1}^{n} e^{z_k}}{\sum_{k=1}^{n} e^{z_k}} - \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}), i \neq j : -p_i(p_j)\right]_{i,j=1,1}^{n,n} \\
&= \left[i = j : p_j(1 - p_i), i \neq j : p_j(-p_i)\right]_{i,j=1,1}^{n,n} \\
&= \left[p_j(I_{i,j} - p_i)\right]_{i,j=1,1}^{n,n}
\end{aligned}
$$

Problem 2

(a)

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{z}} &= \left[\frac{\partial J}{\partial z_j}\right]_{j=1}^{n} \\
&= \frac{-\sum_{i=1}^{n} \partial y_i \ln p_i}{\partial z_j} \\
&= \frac{-\sum_{i=1}^{n} \partial (y_i \ln p_i)}{\partial p_i} * \frac{\partial p_i}{\partial z_j} \\
&= -\sum_{i=1}^{n} \frac{y_i}{p_i} * \frac{\partial p_i}{\partial z_j}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{y_j}{p_j} * \frac{\partial p_j}{\partial z_j} - \sum_{i \neq j}^{n} \frac{y_i}{p_i} * \frac{\partial p_i}{\partial z_j} \\
&= -\frac{y_j}{p_j} * p_j(1 - p_j) - \sum_{i \neq j}^{n} \frac{y_i}{p_i} * -p_i(p_j) \\
&= -y_j(1 - p_j) - \sum_{i \neq j}^{n} -y_i(p_j) \\
&= -y_j + y_j(p_j) + \sum_{i \neq j}^{n} y_i(p_j) \\
&= -y_j + y_j(p_j) + p_j \sum_{i \neq j}^{n} y_i \\
&= -y_j + p_j(y_j + \sum_{i \neq j}^{n} y_i) \\
&= -y_j + p_j \sum_{i=1}^{n} y_i \\
&= -y_j + p_j * 1 = \Big[p_j - y_j \Big]_{j=1}^{n}
\end{aligned}
$$

(b)
```python
class crossEntropyLogit:
    def __init__(self):
        self.p = np.zeros((1, 1))

    def doForward(self, z, y):
        """ Returns J """
        newZ = z - np.max(z, axis=0)
        self.p = np.exp(newZ) / np.sum(np.exp(newZ), axis=0)
        J = -(np.sum(y * np.log(self.p))) / y.shape[1]
        return J

    def doBackward(self, y):
        """ Returns dJ/dz """
        return (self.p - y) / y.shape[1]
```

Problem 3

First iteration:

$$\mathbf{z}^{(0)} = \mathbf{W^{(0)}}\mathbf{X} + \mathbf{b}^{(0)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{p}^{(0)} = \frac{e^{z_i}}{\sum_{i=1}^{2} e^{z_i}} = \begin{bmatrix} \frac{e^0}{e^0+e^0} & \frac{e^0}{e^0+e^0} \\ \frac{e^0}{e^0+e^0} & \frac{e^0}{e^0+e^0} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\begin{aligned} \mathbf{J}^{(0)} &= -\sum_{i=1}^{2} \mathbf{Y} * log(\mathbf{p}^{(0)}) = -\sum_{i=1}^{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} \log(0.5) & \log(0.5) \\ \log(0.5) & \log(0.5) \end{bmatrix} \\ &= -\sum_{i=1}^{2} \begin{bmatrix} \log(0.5) & 0 \\ 0 & \log(0.5) \end{bmatrix} \Rightarrow -log(0.5) \approx 0.6931 \end{aligned}$$

$$d\mathbf{z}^{(0)} = (\mathbf{p^{(0)}} - \mathbf{Y}) \div 2 = \left(\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \div 2 = \begin{bmatrix} \text{-0.25} & 0.25 \\ 0.25 & \text{-0.25} \end{bmatrix}$$

$$d\mathbf{W}^{(0)} = \mathbf{dz}^{(0)} \cdot \mathbf{X}^T = \begin{bmatrix} \text{-0.25} & 0.25 \\ 0.25 & \text{-0.25} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T = \begin{bmatrix} \text{-0.25} & 0.25 \\ 0.25 & \text{-0.25} \end{bmatrix}$$

$$d\mathbf{b}^{(0)} = d\mathbf{z}^{(0)} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \text{-0.25} & 0.25 \\ 0.25 & \text{-0.25} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{W}^{(1)} = \mathbf{W}^{(0)} - d\mathbf{W}^{(0)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \text{-0.25} & 0.25 \\ 0.25 & \text{-0.25} \end{bmatrix} = \begin{bmatrix} 0.25 & \text{-0.25} \\ \text{-0.25} & 0.25 \end{bmatrix}$$

$$\mathbf{b}^{(1)} = \mathbf{b}^{(0)} - d\mathbf{b}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Second iteration:

$$\mathbf{z}^{(1)} = \mathbf{W^{(1)}X} + \mathbf{b}^{(1)} = \begin{bmatrix} 0.25 & \text{-}0.25 \\ \text{-}0.25 & 0.25 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.25 & \text{-}0.25 \\ \text{-}0.25 & 0.25 \end{bmatrix}$$

$$\mathbf{p}^{(1)} = \frac{e^{z_i}}{\sum_{i=1}^{2} e^{z_i}} = \begin{bmatrix} \frac{e^{0.25}}{e^{0.25}+e^{-0.25}} & \frac{e^{-0.25}}{e^{0.25}+e^{-0.25}} \\ \frac{e^{-0.25}}{e^{0.25}+e^{-0.25}} & \frac{e^{0.25}}{e^{0.25}+e^{-0.25}} \end{bmatrix} = \begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix}$$

$$\begin{aligned} \mathbf{J}^{(1)} &= -\sum_{i=1}^{2} \mathbf{Y} * log(\mathbf{p}^{(1)}) = -\sum_{i=1}^{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} \log(0.62) & \log(0.38) \\ \log(0.38) & \log(0.62) \end{bmatrix} \\ &= -\sum_{i=1}^{2} \begin{bmatrix} \log(0.62) & 0 \\ 0 & \log(0.62) \end{bmatrix} \Rightarrow -log(0.62) \approx 0.4741 \end{aligned}$$

$$d\mathbf{z}^{(1)} = (\mathbf{p^{(1)}} - \mathbf{Y}) \div 2 = \left(\begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \div 2 \approx \begin{bmatrix} \text{-}0.19 & 0.19 \\ 0.19 & \text{-}0.19 \end{bmatrix}$$

$$d\mathbf{W}^{(1)} = \mathbf{dz}^{(1)} \cdot \mathbf{X}^T = \begin{bmatrix} \text{-}0.19 & 0.19 \\ 0.19 & \text{-}0.19 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T = \begin{bmatrix} \text{-}0.19 & 0.19 \\ 0.19 & \text{-}0.19 \end{bmatrix}$$

$$d\mathbf{b}^{(1)} = d\mathbf{z}^{(1)} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \text{-}0.19 & 0.19 \\ 0.19 & \text{-}0.19 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{W}^{(2)} = \mathbf{W}^{(1)} - d\mathbf{W}^{(1)} = \begin{bmatrix} 0.25 & \text{-}0.25 \\ \text{-}0.25 & 0.25 \end{bmatrix} - \begin{bmatrix} \text{-}0.19 & 0.19 \\ 0.19 & \text{-}0.19 \end{bmatrix} = \begin{bmatrix} 0.44 & \text{-}0.44 \\ \text{-}0.44 & 0.44 \end{bmatrix}$$

$$\mathbf{b}^{(2)} = \mathbf{b}^{(1)} - d\mathbf{b}^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Problem 4

a) 1 layer, [(10,Linear)] final training accuracy: 93%, final test accuracy: 92%.

b) Final training accuracy: 99%, final test accuracy: 96%.

c) 2 layers, [(140,ReLU),(10,Linear)] final training accuracy:100%, final test accuracy:97%.