The Austalian National University
School of Computing
Dylan Campbell

Semester 1, 2024
Tutorial 5

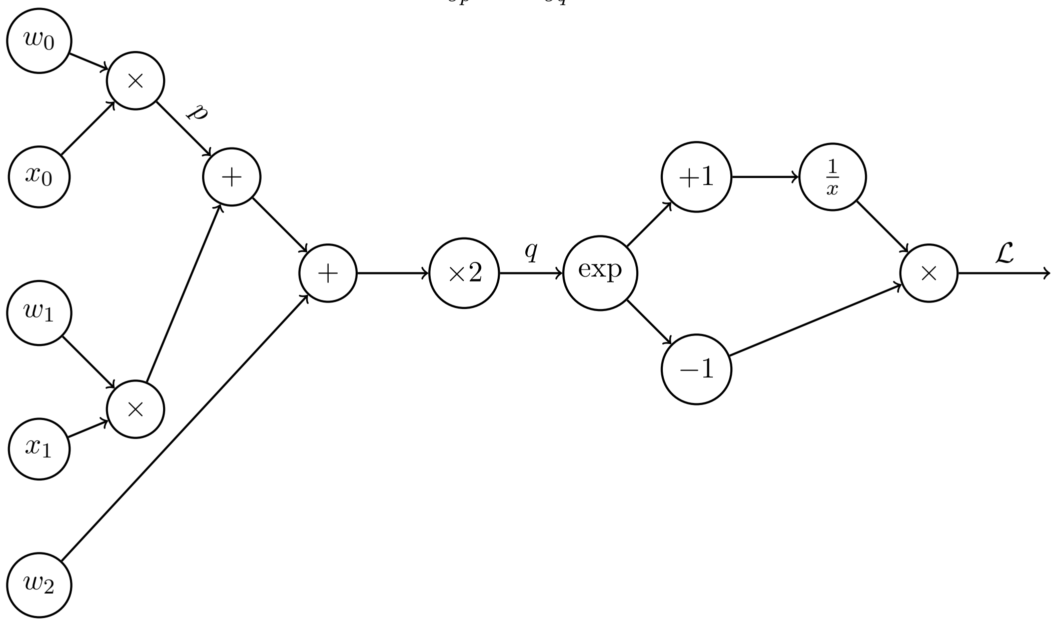
COMP/ENGN 4528/6528: Computer Vision

Question 1 — Matrix Algebra

1. Let $\mathbf{A} = \begin{bmatrix} 2 & 6 \\ 2 & 0 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 2 & 8 \end{bmatrix}$, compute $\mathbf{AB}$.
2. Let $\mathbf{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$, compute $\|\mathbf{x} - \mathbf{y}\|_2$.
3. Let $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$ and $\mathcal{L} = \frac{1}{2}(\mathbf{w}^\top \mathbf{x} - 4)^2$, compute $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$ [1].

Question 2 — Back Propagation

1. Back propagation through the computational graph. The current values are $w_0 = 0.2$, $w_1 = 0.2$, $w_2 = 0.3$, $x_0 = 2$, $x_1 = 3$. p and q define the intermediate variables that are calculated during training, at the specified points in the computation graph. $\mathcal{L}$ is the output of the computational graph. Please provide the gradient $\frac{\partial \mathcal{L}}{\partial p}$ and $\frac{\partial \mathcal{L}}{\partial q}$ based on the back-propagated gradient calculation.



2. Given the linear regression model $\hat{y} = \mathbf{w}^\top \mathbf{x} + b$ and the loss function is defined as $\mathcal{L}(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$. The initial model weights are $\mathbf{w} = \begin{bmatrix} 6 \\ -4 \end{bmatrix}$ and $b = -8$. What is the new model weights after performing one gradient descent step with learning rate 0.01 and training data $\mathbf{x} = \begin{bmatrix} 8 \\ 1 \end{bmatrix}$, $y = 1$.
3. **!!** Given a logistic regression model $\text{Softmax}(\mathbf{Wx} + \mathbf{b})$ where $\mathbf{W} = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 2 & 6 \\ -1 & 1 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -2 \\ -1 \\ -2 \end{bmatrix}$.

   The Softmax function is defined as $\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$. For the training data, you have

[1] read `https://en.wikipedia.org/wiki/Matrix_calculus` for more matrix calculus contents

$\mathbf{x} = \begin{bmatrix} -2 \\ -5 \\ 4 \end{bmatrix}$ and the ground truth label $\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Calculate the gradient of the cross-entropy loss ($\mathcal{L} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$)[2] with respect to the bias vector $\mathbf{b}$.

[2] The meaning of each variable is mentioned in `https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html#cross-entropy`