50.039 – Theory and Practice of Deep learning

Alex

Week 03

[The following notes are compiled from various sources such as textbooks, lecture materials, Web resources and are shared for academic purposes only, intended for use by students registered for a specific course. In the interest of brevity, every source is not cited. The compiler of these notes gratefully acknowledges all such sources.]

Due: week4 Wednesday, 6pm

1 Task 1:

Consider the loss:

$$(-1) \cdot \sum_{i=1}^{n} y_i \log\left(h(x_i)\right) + (1-y_i)\log\left(1-h(x_i)\right)$$

Prove that The gradient with respect to w will be:

$$\nabla_w L = \nabla_w \left((-1) \cdot \sum_{i=1}^{n} y_i \log(s(w \cdot x_i)) + (1-y_i)\log\left(1-s(w \cdot x_i)\right) \right)$$

$$\nabla_w L = \sum_{i=1}^{n} x_i (s(w \cdot x_i) - y_i) = \sum_{i=1}^{n} x_i (h(x_i) - y_i)$$

You can use

$$\frac{\partial log(s(a))}{\partial a} = 1 - s(a)$$

$$\frac{\partial log(1-s(a))}{\partial a} = -s(a)$$

and we know $h(x) = s(w \cdot x)$

2 Task 2:

Which einsum notation is required to implement the following operations? Remember it is a pair

$$indices_1, indices_2, indices_3, \ldots, indices_n - > indices_r, [t_1, t_2, t_3, \ldots, t_n]$$

- $$C_{j,k} = \sum_i A_{ijk} b_i$$

- $$C_j = \sum_{i,k} A_{ijk} b_{ik}$$

- $$A_{ik} = \sum_{j,l} A_{ijkl}$$

- yes this is not the same as before, note the change in index ordering
  $$A_{ki} = \sum_{j,l} A_{ijkl}$$

- $$C_i = \sum_{j,k} A_{ijk} A_{ijk}$$

- $$C = x^\top Ax, x \in \mathbb{R}^d,\ 1 - tensor, A \in \mathbb{R}^{d\times d},\ 2 - tensor,$$

- $$C = AG^\top B, A \in \mathbb{R}^{d\times e},\ 2 - tensor, G \in \mathbb{R}^{f\times e},\ 2 - tensor, B \in \mathbb{R}^{f\times l},\ 2 - tensor,$$

  The result is a tensor of what order here ? in any case there is more than one possible output index ordering in the sense of C_{ijk} vs C_{jki} vs C_{kij} and so on . any output index ordering is okay here

- $$C_{????} = \sum_{cd} A_{abcd} B_{bcde} E_{cdef}$$

  any output index ordering is okay here again

3 Task 3:

which of these pairs of tensor shapes are broadcastable ? If they are, what is the final result shape ?

- $$\begin{array}{r} (3,1,2,3) \\ (5,3) \end{array}$$

- $$(3,2,1,3,4)$$
  $$(5,3,4)$$

- $$(3,2,1,3,4)$$
  $$(5,1,4)$$

- $$(3,2,1,3,2)$$
  $$(5,3,1)$$

- $$(3,2,1,3,2)$$
  $$(1,3,1,2)$$

- $$(7,1)$$
  $$(7)$$