Theory and Practice of Deep Learning

Theory HW #03

Nigel Chan 1002027

Task 1 : Some Einsum

a) $C_{j,k} = \sum_{i} A_{ijk} b_i$ → einsum('ijk,i->jk', A, b)

b) $C_j = \sum_{i,k} A_{ijk} b_{ij}$ → einsum('ijk,ik->j', A, b)

c) $A_{ik} = \sum_{j,l} A_{ijkl}$ → einsum('ijkl->ik', A)

d) $A_{ki} = \sum_{j,l} A_{ijkl}$ → einsum('ijkl->ki', A)

e) $C_i = \sum_{j,k} A_{ijk} A_{ijk}$ → einsum('ijk,ijk->i', A, A)

f) $C = x^T A x$ → einsum('i,ij,j->', x, A, x)

g) $C = AG^T B$ → einsum('ik,lk,lj->ij', A, G, B) , second order

h) $C_{????} = \sum_{cd} A_{abcd} B_{bcde} E_{cdef}$ → einsum('abcd,bcde,cdef->abef', A, B, E)

Task 2

$$P(X^{(1)} < 0 \mid Y = -1) = 0.5$$
$$P(X^{(1)} < 0 \mid Y = +1) = 0.5$$

$$f_0(x) = 2\,\mathbb{I}[x^{(1)} \geq 0] - 1 = \begin{cases} -1 & x^{(1)} < 0 \\ +1 & x^{(1)} \geq 0 \end{cases}$$

a) $\cancel{E_{(x,y)\sim p}[\mathbb{I}[f_0(x) \neq y]] = P(f_0(x)=1, y=-1) + P(f_0(x)=-1, y=1)}$

using the formula,
$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

$$\begin{aligned} E_{(x,y)\sim p}[\mathbb{I}[f_0(x) \neq y]] &= P(f_0(x)=1, y=-1) + P(f_0(x)=-1, y=1) \\ &= P(f_0(x)=1 \mid y=-1)P(y=-1) + P(f_0(x)=-1 \mid y=1)P(y=1) \\ &= P(X \geq 0 \mid y=-1)(1-0.5) + P(X<0 \mid y=1)(0.5) \\ &= (1-0.5)(0.5) + 0.5(0.5) \\ &= 1(0.5) \\ &= 0.5 \end{aligned}$$

b) For the points with class -1, the probability that the error on the training dataset being zero is $P(f_0(x_i) = -1 \mid y_i = -1) = 0.5$, based on the resulting probability we found in part (a). Likewise for class $+1$ points, $P(f_0(x_i) = 1 \mid y_i = 1) = 0.5$

$$\begin{aligned} \text{Overall probability} &= P(\text{first } N/2 \text{ points of class } -1) \times P(\text{last } N/2 \text{ points of class } +1) \\ &= [P(f_0(x_i) = -1 \mid y_i = -1)]^{N/2} \times [P(f_0(x_i) = 1 \mid y_i = 1)]^{N/2} \\ &= (0.5)^{N/2} \times (0.5)^{N/2} \\ &= 0.5^N \end{aligned}$$

c) We use Binomial Distribution here, where $n = D$, $p = 0.5^N$, $k = K$

d) Probability to draw N samples with at least one dimension d out of D dimensions

$$\begin{aligned} &= [P(X \geq 1)]^N \\ &= [1 - P(X<1)]^N \\ &= [1 - P(X=0)]^N \\ &= \left[1 - {}^{D}C_{0}\left(\tfrac{1}{2}\right)^0\left(\tfrac{1}{2}\right)^{D-0}\right]^N \\ &= \left[1 - \left(\tfrac{1}{2}\right)^D\right]^N \end{aligned}$$

e) As $D \to \infty$, $\left[1 - \left(\tfrac{1}{2}\right)\right]^N \to 1$ $\therefore$ probability of zero error $\to 1$

The complexity for convergence given N samples is an exponential of the form : $O\left(\tfrac{1}{2}\right)^{D^N}$