Chia Yew Ken 1002675

1. Some Einsum
    a. ijk,i -> jk
    b. ijk,ik -> j
    c. ijkl -> ik
    d. ijkl -> ki
    e. ijk,ijk -> i
    f. d,dd,d -> dd
    g. de,ef,fl -> dl (2-tensor)
    h. abcd,bcde,cdef -> af
2. Overfitting with more and more dimensions
    a. compute $E_{(x,y)\sim P}[I[f_0(x) \neq y]]$. Show your work in detail. This works for any value of $P(Y = +1)$.
        i. p(x|y) = p(x,y)p(y)
        ii. If y = -1, f(x) = +1/-1 (50/50), E[f(x) != y] = 0.5
        iii. If y = +1, f(x) = +1,-1 (50/50), E[f(x) != y] = 0.5
        iv. Thus E = 0.5 regardless of p(y=+1)
    b. What is the probability that we draw N samples such that the error on this training dataset is zero under $f_0(x)$? Express this event in terms of conditions to x_i for the first $N/2$ points and for the last $N/2$ points. Then compute its probability under above $P(X|Y)$.
        i. First N/2 points, y = -1
            1. p(f(x) == y) = 0.5 for each sample
            2. p(error == 0) = p(f(x) == -1) ^ (N/2) = 0.5 ^ (N/2)
        ii. Last N/2 points, p(y=+1) unknown
            1. For each sample
                a. If y = -1, p(f(x) == y) = 0.5
                b. Sample for (y = +1)
                c. Thus p(f(x) == y) = 0.5
            2. p(error == 0) = same for first N/2 points = 0.5 ^ (N/2)
        iii. Thus overall p(error == 0) = 0.5 ^ (N/2 + N/2) = 0.5 ^ N
    c. What is the probability distribution that we draw N samples such that in exactly K out of D dimensions (remember $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(D)})$) $\{d_1, \ldots, d_K\} \subset 1, \ldots, D$ $f_0(x^{(d_k)})$ achieves zero training error? Give its name and its parameters.
        i. In 1 dimension, for N samples, p(error == 0) = 0.5 ^ N
        ii. Common distributions
            1. Bernoulli is binary, single trial
            2. Uniform is multiple outcomes, single trial
            3. Binomial is multiple Bernoulli trials
            4. Normal is bell-shaped, continuous

5. Poisson is number of events over time
6. Exponential is time interval between events

iii. We can treat each dimension as a trial, outcome is whether error over N samples is zero, which is binary. Each dimension is statistically independent. Hence we can model with a Binomial distribution with formula:

1. $$P(x) = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$, where
2. n = D total dimensions,
3. x = K successes (dimensions where error == 0)
4. p = P(success) = 0.5 ^ N samples
5. q = P(failure) = 1 - P(success) = 1 - (0.5^N)

d. What is the precise probability that we draw N samples such that in at least one dimension d out of D dimensions $f_0(x^{(d)})$ achieves zero training error?

i. From Binomial framing above, this case is e quivalent to 1 - P(x=0)
ii. Thus probability = 1 - (D!/D! * p^0 * q^D) = 1 - q^D = 1 - (1-0.5^N)^D

e. What is the limit of this probability as $D \to \infty$? What is the $\mathcal{O}(\cdot)$ complexity of the convergence of this limit as a function of D ?

i. q = (1-0.5^N) is always < 1, assuming N >= 0
ii. As D -> infinity, q -> 0, thus limit is 1 - 0 = 1
iii. The convergence towards 1 is O(1 - q^D), where q = (1-0.5^N) is in range [0, 1]