Homework #0 - B

Spring 2020, CSE 446/546: Machine Learning
Richy Yun
Due: 4/8/19 11:59 PM

Probability and Statistics

B.1 *[1 points]* To calculate the expected value, we want to find the CDF then calculate the PDF:

$$CDF_Y = p[\max\{X_1, \ldots, X_n\} \leq x]$$

Since the maximum means everything is less than or equal to, and all X are independent and identically distribtued, we can rewrite the above as:

$$\begin{aligned} CDF_Y &= p[X_1, \ldots, X_n \leq x] \\ &= p[X_1 \leq x]p[X_2 \leq x] \ldots p[X_n \leq x] \\ &= \prod_{i=1}^{n} p[X_i \leq x] \\ &= (p[X_i \leq x])^n \end{aligned}$$

The PDF is then the derivative of the CDF:

$$PDF_Y = \frac{d}{dn}(p[X_i \leq x])^n = n(p[X_i \leq x])^{n-1}$$

$p[X_i \leq x]$ is the CDF of X_i. Since they are a uniform distribution from 0 to 1:

$$\begin{aligned} PDF_{X_i} &= \frac{1}{b-a} = \frac{1}{1-0} = 1 \\ CDF_{X_i} &= \int_0^x 1dt = x \end{aligned}$$

Now we can calculate the expected value using the definition, from 0 to 1 as that is the limit of the distribution:

$$\begin{aligned} \mathbb{E}[Y] = \int_0^1 x \times PDF_Y dx = \int_0^1 x \times n(x)^{n-1} dx &= \int_0^1 n(x)^n dx \\ &= n\frac{x^{n+1}}{n+1} \qquad \text{for } x = 0 \text{ to } 1 \\ &= \frac{n}{n+1} \end{aligned}$$

Linear Algebra and Vector Calculus

B.2 *[1 points]* The definition of matrix multiplication provides:

$$AB_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

Therefore, for the diagonals we have:

$$AB_{ii} = \sum_{k=1}^{n} A_{ik} B_{ki}$$

$$BA_{ii} = \sum_{k=1}^{n} B_{ik} A_{ki}$$

We can rewrite the traces as:

$$\mathrm{Tr}(AB) = \sum_{i} ABii = \sum_{i=1}^{m} \sum_{k=1}^{n} A_{ik} B_{ki}$$

$$\mathrm{Tr}(BA) = \sum_{i} BAii = \sum_{i=1}^{n} \sum_{k=1}^{m} B_{ik} A_{ki}$$

Now, simply by rearranging using the commutative property:

$$\mathrm{Tr}(AB) = \sum_{i=1}^{m} \sum_{k=1}^{n} A_{ik} B_{ki} = \sum_{i=1}^{n} \sum_{k=1}^{m} B_{ik} A_{ki} = \mathrm{Tr}(BA) =$$

B.3 *[1 points]*

a. In the case v_i has a size of $n \times 1$ the result of the product is a matrix of size $n \times n$. However, as the results of each row element is dependent on the first matrix, the product has columns that are linearly dependent with a rank of 1. Summing n matrices of rank 1 results in a matrix with rank n when $n < d$ and rank d otherwise.

b. The minimum possible rank is 1 as each v_i could be a linear combination of one another. The maximum possible is n as long as $n < d$ and d otherwise, as each vector could be linearly independent of one another, but rank cannot be larger than the smallest dimension.

c. First we need to determine the minimum and maximum possible rank of matrix A. The minimum rank is 1 as each column could be linear combinations of one column, and the maximum rank is d since $D > d$. Thus, the product Av_i also has a minimum and maximum rank of 1 and d respectively. Although the result of $(Av_i)(Av_i^T)$ is a matrix of size $D \times D$ since Av_i has maximum rank d the resulting product has maximum rank d similar to part (a). However, summing the matrices results in the sum of the ranks, so we end up with a matrix with a minimum rank n assuming $n < D$ and a maximum rank D.

d. The rank of AV depends on A as mentioned in part (a). Therefore, regardless of the rank of V the minimum and maximum rank of AV is 1 and d respectively.