

ESL Chapter 2 Exercises

Chris Regan

July 9, 2023

Abstract

This covers ESL exercises I have done for Ch. 2 (p.39-41 of the book)

1 Ex. 2.3

INCOMPLETE – not solved yet.

Let us define (R_1, \dots, R_N) as a set of i.i.d. random variables which are drawn from a uniform distribution inside a p -dimensional unit hyper-sphere. The expected value of the first order statistic $R_{(1)}$ is the quantity we are interested in (since it represents the expected minimum value from the sample of size N).

The CDF of the first order statistic is given by $F_{R_{(1)}}(r) = 1 - [1 - F_R(r)]^n$, so we first need to find the CDF of a uniform distribution inside of a p -dimensional unit hyper-sphere.

For $0 < r < 1$, we know that the PDF is a constant, so we first need to find the normalizing constant (and zero for other values of r), and we know that the normalizing constant must be the volume of the hyper-sphere (which we call V_p).

If we work with hyper-spherical coordinates, such that r is the distance from the origin and $\vec{\Omega} = (\theta_1, \dots, \theta_{p-1})$ is our vector of angular coordinates, then we know the PDF is $f(r, \vec{\Omega}) = \frac{1}{V_p}$ for $0 < r < 1$ and zero otherwise. We want to integrate out $\vec{\Omega}$. We know that d^pV in Cartesian coordinates would become $r^{p-1}drd\vec{\Omega}$ when transforming to hyper-spherical coordinates. We first want to find $\int d\vec{\Omega}$ integrated over all angular positions in the sphere, leveraging the fact that the integral over the entirety of the sphere would be equal to V_p

$$\begin{aligned} V_p &= \int d^pV = \\ &= \int \int r^{p-1}drd\vec{\Omega} = V_p \\ &= \left[\int_0^1 r^{p-1}dr \right] \cdot \left[\int d\vec{\Omega} \right] \\ &= \frac{1}{p} \cdot \int d\vec{\Omega} \end{aligned} \tag{1}$$

Thus, $\int d\vec{\Omega} = p \cdot V_p$ when integrated over all possible angular positions given that $0 < r < 1$. We can use this to first compute the marginal CDF that only depends on r , and then differentiate that to get the marginal PDF.

$$\begin{aligned}
F_R(r) &= \int f_R(r, \vec{\Omega}) r^{p-1} d\vec{\Omega} \\
&= \frac{1}{V_p} \cdot \int d\vec{\Omega} \int_0^r \rho^{p-1} d\rho \\
&= \frac{1}{V_p} \cdot p \cdot V_p \int_0^r \rho^{p-1} d\rho \\
&= \frac{1}{V_p} \cdot p \cdot V_p \cdot \left[\frac{\rho^p}{p} \right]_{\rho=0}^{\rho=r} \\
&= r^p
\end{aligned} \tag{2}$$

Thus, the CDF of the first order statistic is $F_{R_{(1)}} = 1 - [1 - F_R(r)]^n = 1 - [1 - r^p]^n$, so the PDF of the first order statistic is $f_{R_{(1)}}(r) = \frac{d}{dr} F_{R_{(1)}}(r) = -n[1 - r^p]^{n-1}(-pr^{p-1}) = np[1 - r^p]^{n-1}r^{p-1}$. Thus, to get the average distance from the origin, we need to solve

$$\begin{aligned}
E(R) &= np \int_0^1 r \cdot [1 - r^p]^{n-1} r^{p-1} dr \\
&= np \int_0^1 r^p [1 - r^p]^{n-1} dr
\end{aligned} \tag{3}$$

ISSUE IS NOT BEING ABLE TO SOLVE THIS INTEGRAL. Based on numerical tests in python, this seems to be the right integral to solve, but not sure how to do it.

2 Ex. 2.7

2.1 (a)

For linear regression, we first solve the general case where each of the N observations \vec{x}_i is a $p \times 1$ vector (so χ is a $N \times p$ matrix and \vec{y} is a $N \times 1$ vector of responses). We know that the linear regression estimate is $\hat{f}(x_0) = \hat{\beta}^T \vec{x}_0$ where $\hat{\beta} = (\chi^T \chi)^{-1} \chi^T \vec{y}$. Thus, $\hat{f}(\text{vec} x_0) = \vec{y}^T \chi (\chi^T \chi)^{-1} \vec{x}_0$. We can alternatively express this as a summation, and we get $\hat{f}(\vec{x}_0) = \sum_{i=0}^N y_i \cdot \vec{x}_i^T (\chi^T \chi)^{-1} \vec{x}_0$, and we can see that our weights are $l_i(\vec{x}_0; \chi) = \vec{x}_i^T (\chi^T \chi)^{-1} \vec{x}_0$. In the case where we only have one parameter (i.e. x_i is a scalar), this becomes $l_i(x_0; \chi) = \frac{x_i x_0}{\sum_{i=0}^N x_i^2}$ (assuming it's fit without an intercept term).

For k -nearest neighbor regression, we have $\hat{f}(x_0) = \frac{1}{k} \sum_{x_i \in N_k(x_0)} y_i$, so $l_i(x_0; \chi) = \frac{1}{k}$ if $x_i \in N_k(x_0)$ and zero otherwise.