Elements of Machine Learning
Assignment 3

William LaCroix (7038732)
wila00001@stud.uni-saarland.de

Problem 2 (T, 6 Points). **The Bootstrap.**

We will now derive the probability that a given observation is part of a bootstrap sample of size n. Suppose that we obtain a bootstrap sample from a set of n observations.

1. [2pts] What is the probability that the first bootstrap observation is not the jth observation from the original sample? Justify your answer.

From a training sample Z of size n, the first bootstrap observation b_1 will have a $\frac{1}{n}$ chance of being any given z_i, and a $\frac{n-1}{n} = \left(1 - \frac{1}{n}\right)$ chance of not being a specific z_j. $P\left(b_1 = z_j\right) = \left(1 - \frac{1}{n}\right)$

2. [2pts] Argue that the probability that the jth observation is not in the bootstrap sample is $\left(1 - \frac{1}{n}\right)^n$

As in part 1, where the probability of a single observation b_i not being the jth training observation is given by $P\left(b_i \neq z_j\right) = \left(1 - \frac{1}{n}\right)$, the joint probability of z_j not appearing a given bootstrap set Z^{*B} is given by: $P\left(z_j \notin Z^{*B}\right) = \Pi_{i=1}^{n}\left(1 - \frac{1}{n}\right) = \left(1 - \frac{1}{n}\right)^n$

3. [2pts] Comment on the behavior of the above probabilities with increasing sample size n.

As sample size n increases, the individual probability (part 1) of not including a given observation approaches 1, that is, $P\left(b_i \neq z_j\right) \frac{lim}{n\to\infty} = 1$, and by extension, the joint probability (part 2) of z_j not appearing in bootstrap set Z^{*B}, $P\left(z_j \notin Z^{*B}\right) = \left(1 - \frac{1}{n}\right)^n \frac{lim}{n\to\infty} = 1$, would imply that for sufficiently large n, in any given Z^{*B}, any one z_j being in the bootstrap set is impossible, since $P\left(z_j \in Z^{*B}\right) = 1 - \left(1 - \frac{1}{n}\right)^n \frac{lim}{n\to\infty} = 0$. Which is not a particularly useful feature, and frankly feels wrong.