Cover page for answers.pdf
CSE353 Fall 2020 - Machine Learning - Homework 1

Your Name: Piotr Milewski

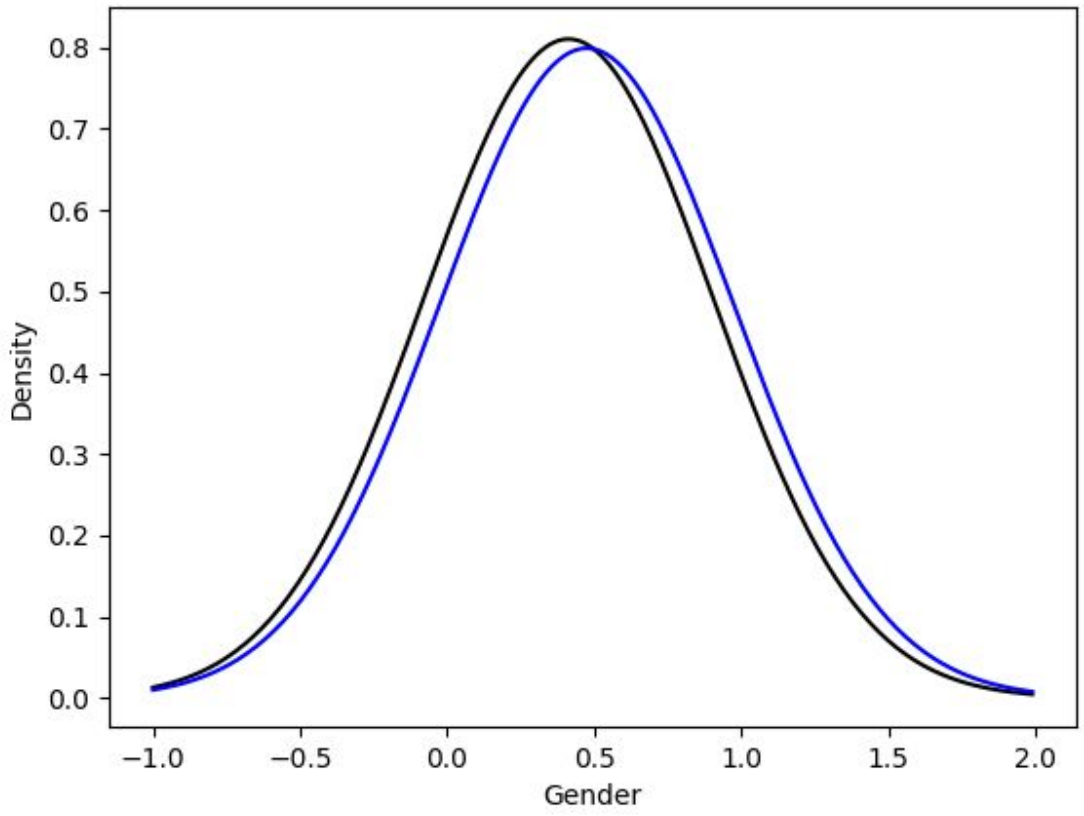
Solar ID: 112291666

NetID email address: piotr.milewski@stonybrook.edu

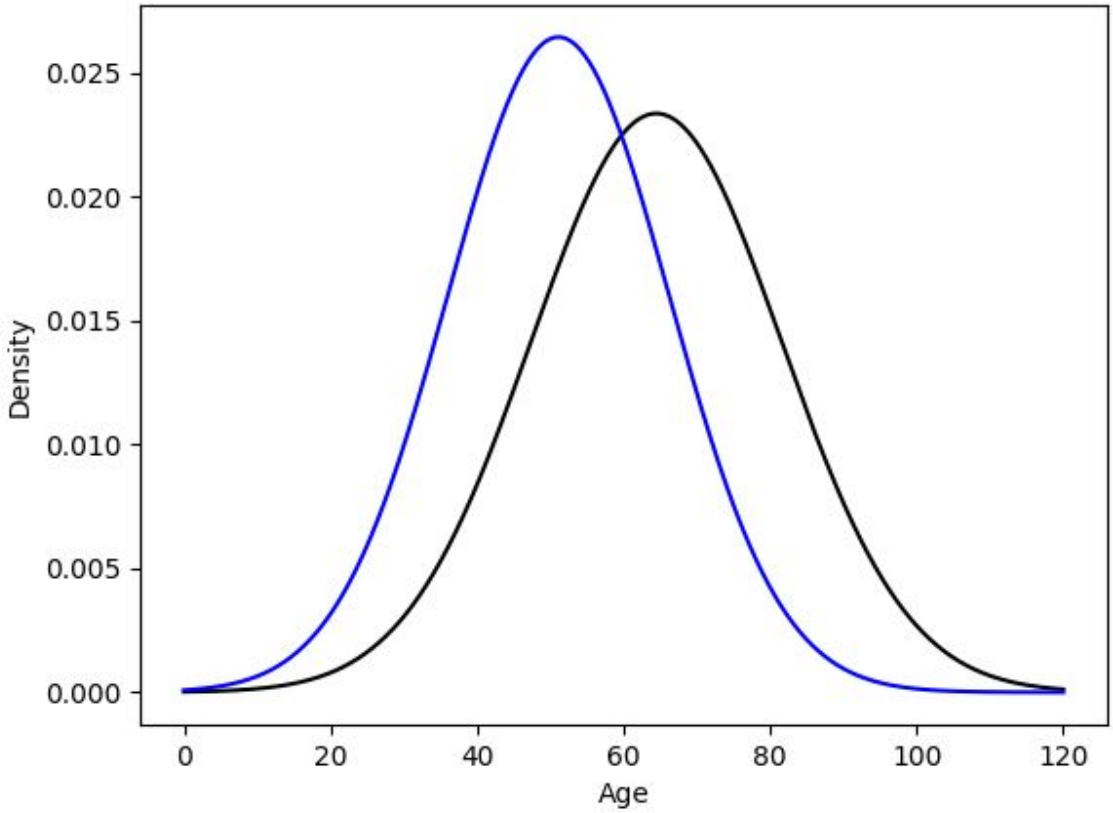
Names of people whom you discussed the homework with:

Question 1.2

a. mu0: 64.45652174, 0.41304348
   var0: 291.856805, 0.242438563
   mu1: 51.10055866, 0.47486034
   var1: 227.76642427, 0.249368
b. Feature: Gender



Feature: Age

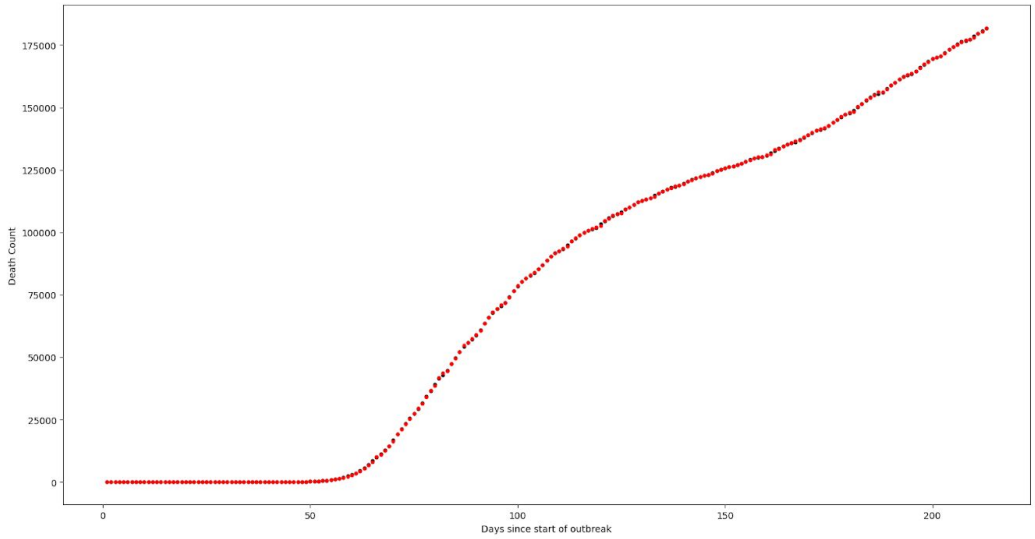


c. It is not a good idea to approximate gender by a Gaussian distribution because the Gaussian distributions for gender for where the patient survived or didn't survive are almost identical in comparison to the Gaussian distributions for age. We can therefore determine from the two graphs that gender isn't an important factor to consider when determining whether or not a patient who has COVID will survive.

Question 2.2

a. w1: 3.32971812e-03
w2: -5.97868846e-03
w3: -1.07181360e-02
w4: 2.97430033e-02
w5: -1.33508726e-02
w6: -2.94346183e-03
w7: -8.47872012e-05
w8: 1.74178344e+00
w9: -1.06486597e+00

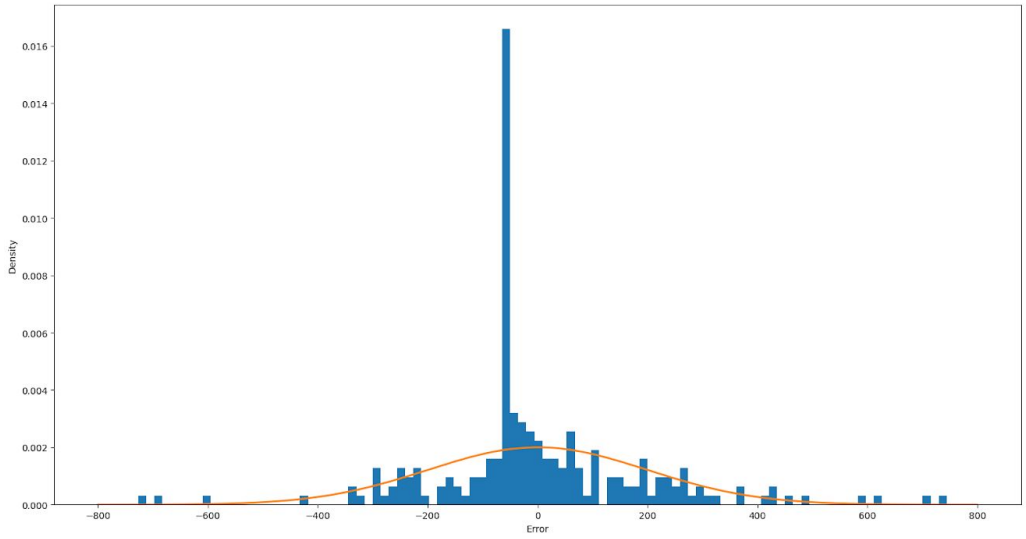
w10: 6.02420242e-01
w11: -3.68951686e-01
w12: -9.27692517e-02
w13: 6.78640538e-01
w14:-4.98447204e-01
b: 58.15728936622327



b.

If you would like to zoom in here is an imgur link: https://imgur.com/a/tmnwn3o

c. Mean of this Gaussian: 8.771499958295952e-12
Variance of this Gaussian: 39761.561252769374



d.

If you would like to zoom in here is an imgur link: https://imgur.com/a/looVU4m
I think Gaussian is a good approximation for the distribution of the errors since as we can see above most of the errors are distributed along the Gaussian function with an outlier at -58 which is due to the death count at the beginning of the outbreak staying at 0.