Assignment 1

Deep Learning, WS22

Team Members		
Last name	First name	Matriculation Number
Amering	Richard	1331945

1 Task 1 - Maximum Likelihood Estimation

1.1 Derivation

Likelihood of a single sample of dimension m:

$$p(x|\theta) = \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

Likelihood for a set X of N samples:

$$X = \{x_1, x_2..x_N\}$$

$$p(X|\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$$

$$|\Sigma| = \sigma^6$$

Add logarithms to both sides of the equation to obtain log-likelihood:

$$log(p(X|\theta)) = \sum_{i=1}^{N} log(\frac{1}{\sqrt{(2\pi)^3|\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)})$$

$$log(p(X|\theta)) = \sum_{i=1}^{N} log(\frac{1}{\sqrt{(2\pi)^3|\Sigma|}}) + log(e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)})$$

$$log(p(X|\theta)) = \sum_{i=1}^{N} log(\frac{1}{\sqrt{(2\pi)^3|\Sigma|}}) - \frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

$$(x_i - \mu)^T \Sigma^{-1}(x_i - \mu) = (x_i - \mu)^T \sigma^{-2} I^{-1}(x_i - \mu) = \sigma^{-2}(x_i - \mu)^T (x_i - \mu)$$

Inner product given by:

$$(x_i - \mu)^T (x_i - \mu) = \sum_{j=1}^{m} (x_{i,j} - \mu_{i,j})^2$$

1.2 Finding the Maximum

For finding the maximum likelihood estimation of μ, the log-likelihood is derived wrt. the different vector components of μ, and then set to zero. Since the log-expression is monotonically increasing and without upper bound, a root in the derivative of the expression must be corresponding to a minimum in the log function, and therefore a minimum in the original likelihood function. This is because a monotonic function without upper bounds like the likelihood function does not have maxima. The addition of logarithms does not

alter the function in terms of locations of its extrema. In the following, the derivation is carried out for component j, which is one component of the m-dimensional data.

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \frac{d}{d\mu_j}(x_{i,j} - \mu_j)^2$$

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \frac{d}{d\mu_j}(x_{i,j}^2 - 2x_{i,j}\mu_j + \mu_j^2)$$

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{\sigma^2} \sum_{i=1}^{N}(-x_{i,j} + \mu_j)$$

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{\sigma^2} \sum_{i=1}^{N}(-x_{i,j}) + N\mu_j \stackrel{!}{=} 0$$

Finally giving the maximum likelihood estimation for μ, which is just the mean value of component j of samples x_i in the set:

$$\mu_j = \frac{1}{N} \sum_{i=1}^{N}(x_{i,j})$$

2 Task 2 - Practical: Selecting data and coding

2.1 Selecting data

The first task of the practical is to choose a dataset, and from that choose 3 metrics that are reasonably close to be normal distributed. The chosen dataset is 'social_capital_zip.csv'.

To make a quick assesment about the distribution of the underlying distribution of different metrics provided in the raw data, histogram plots are created. These plots show the dnumber of datapoints in a umber of bins, after the raw data has been binned according to lower and upper bound of the individual bins. There are many metrics in the raw data that do not at all follow a gaussian distribution. Many metrics, most notably the pop_2018 metric, seem to follow Benfords law (as we should expect), which is very much skewed towards low population values, shown in figure 1
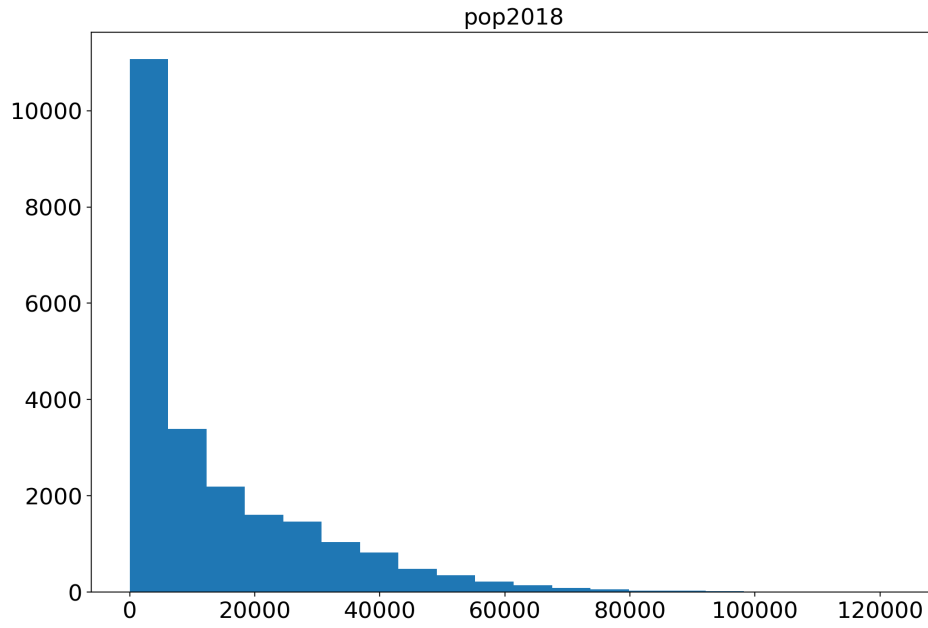
Figure 1: Population in the year 2018 appears to follow Benford's law

After a bit of trial and error, the metrics "exposure_grp_mem_zip", "nbhd_ec_zip" and "bias_grp_mem_zip" were chosen, with their distributions shown in figure 2, 3 and 4:
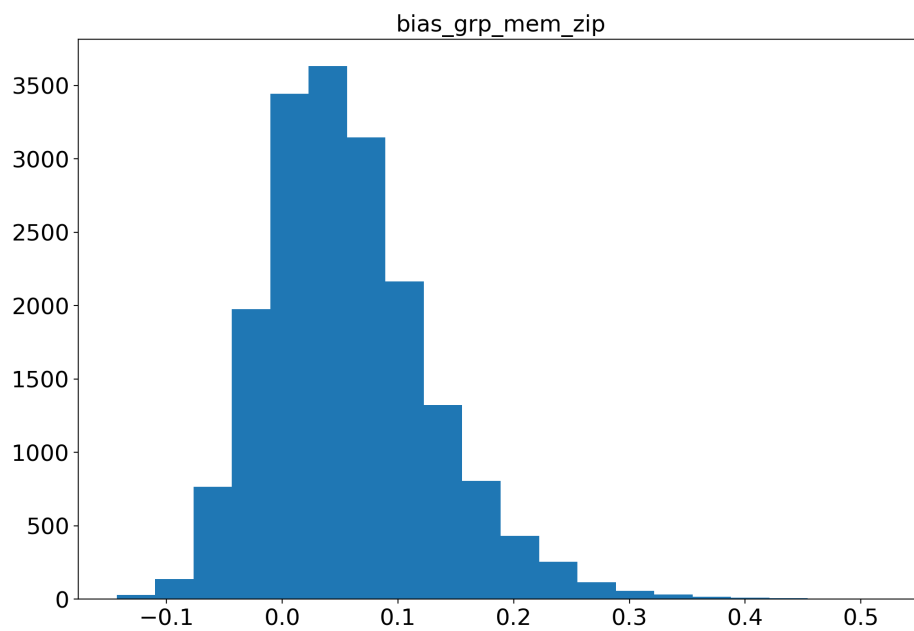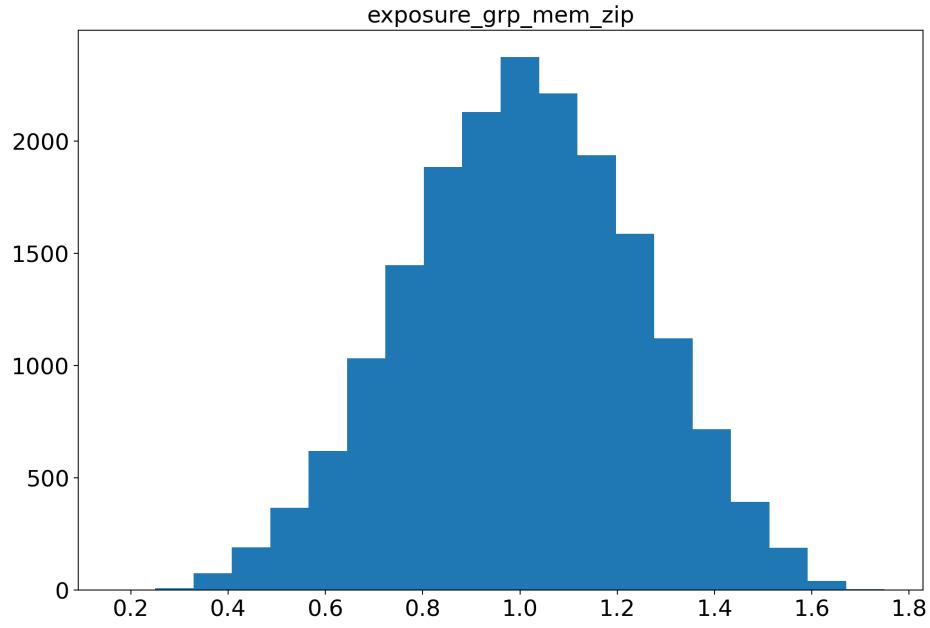

Figure 2: Size of bias groups

Figure 3: Mean exposure to high-SES individuals by ZIP code for low-SES individuals
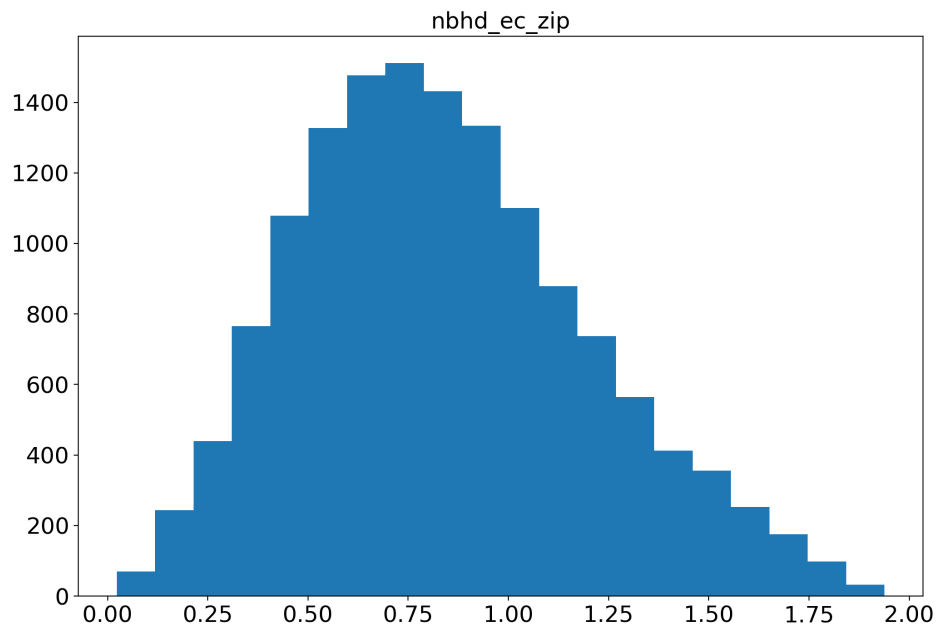

Figure 4: ec grp mem zip divided by exposure grp mem zip, all subtracted from one

To test if the three individual dataseries combine to a 3D gaussian distribution, we create a 3D scatter-plot and assess the qualitative appearance in figure 5:
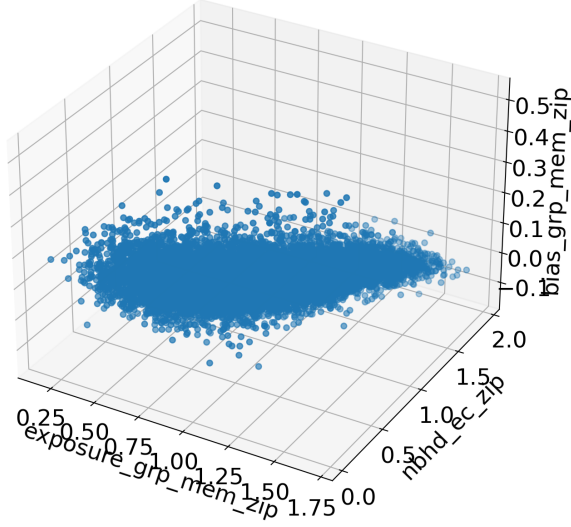
4

Figure 5: 3D scatter plot of the chosen metrics

The 3D scatter plot creates a point cloud in 3D space that appears to be reasonably well distributed without much skewing towards one area or without exhibiting multiple clusters. The chosen metrics appear to satisfy the requirements given in the task description.

2.2 Calculating the maximum likelihood estimation of μ

Finally, the maximum likelihood estimation is calculated as the mean of the three dataseries, giving the vector:

$$\mu = \begin{bmatrix} \mu_{egm} \\ \mu_{nbhd\_ec} \\ \mu_{bgm} \end{bmatrix} = \begin{bmatrix} 1.005 \\ 0.839 \\ 0.057 \end{bmatrix}$$