Assignment 4: CS 215

Devansh Jain	Harshit Varma
190100044	190100055

November 13, 2020

Question 2

Instructions for running the code:

1. Unzip and `cd` to `q2/code`, under this find the file named `q2.m`
2. Run the file, required plots will be generated and saved to the `q2/results` folder
   `d_x.jpg`: scatter plots with the required lines for each N
   `mean_err.jpg` and `cov_err.jpg`: box plots of error measure of mean and covariance

(a)

Let the 2D Gaussian Random Vector be $X = AW + \mu$, where W is a vector of 2 i.i.d Standard Normal Random Variables.
We are given $\mu = [1\ 2]^T$, we need to find A from covariance matrix $C = AA^T$.
C is a Symmetric Positive Definite Matrix, and can be represented as QDQ^T, where Q is a orthogonal Matrix and D is a Diagonal Matrix. These matrices can easily be calculated by `eig` function in `MATLAB`.
$C = QDQ^T = AA^T \implies A = QD^{0.5}$,
D is a diagonal matrix with positive diagonal elements as Eigenvalues of a SPD Matrix are positive.

```
mu = [1; 2];
cov = [1.6250 -1.9486; -1.9486 3.8750];
[Q, D] = eig(cov);
A = Q*sqrt(D); % square root of a diagonal matrix is square root of individual elements
```

We can get values from i.i.d Standard normal distribution using `randn`, this would constitute the W.

```
x = A * randn(2,N) + mu; % values from X
```

Here, N is size of the sample data.
The sample points from the 2D Gaussian are stored in variable `x` in the workspace in form of a $2 \times N$ matrix

(b)

We proved in class that ML Estimate $\hat{\mu}$ is sample mean $\sum_{i=1}^{N} x_i/N$.
It can be computed in `MATLAB` using `sum`.

```
mean_mle = sum(x,2) / N;
```

For calculating the error measure, I defined `norm` function to calculated 2-norm for vector and Frobenius norm for matrices (required in (c)).

```
function ret = norm(x)
    ret = sqrt(sum(x.*x, 'all'));
    % 'all' makes this function useful for calculating 2-norm for vectors
    % and frobenius norm for matrices
end
```

The error measure for mean is calculated and stored in a 100×5 matrix `mean_err`.

```
mean_err(iter,lN) = norm(mu - mean_mle) / norm_mu; % Where lN = log10(N)
```

(c)

We proved in class that ML Estimate $\hat{C}_N$ is sample covariance $\sum_{i=1}^{N}(x_i - \hat{\mu})(x_i - \hat{\mu})^T/N$.
It can be computed in `x_mu` $\equiv X - \mu$ which was created while generating the random variables.

```
cov_mle = (x-mean_mle)*(x-mean_mle)' / N;
```

The error measure for covariance is calculated and stored in a 100×5 matrix `cov_err`.

```
cov_err(iter,lN) = norm(cov - cov_mle) / norm_cov; % Where lN = log10(N)
```

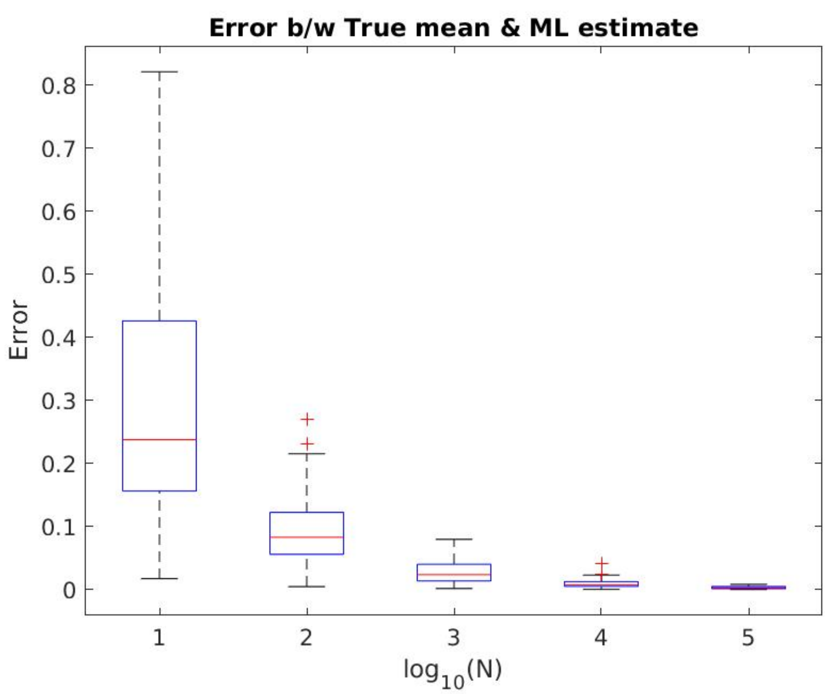

Figure 1: Box plot of the error between the true mean μ and the ML estimate $\hat{\mu}_N$

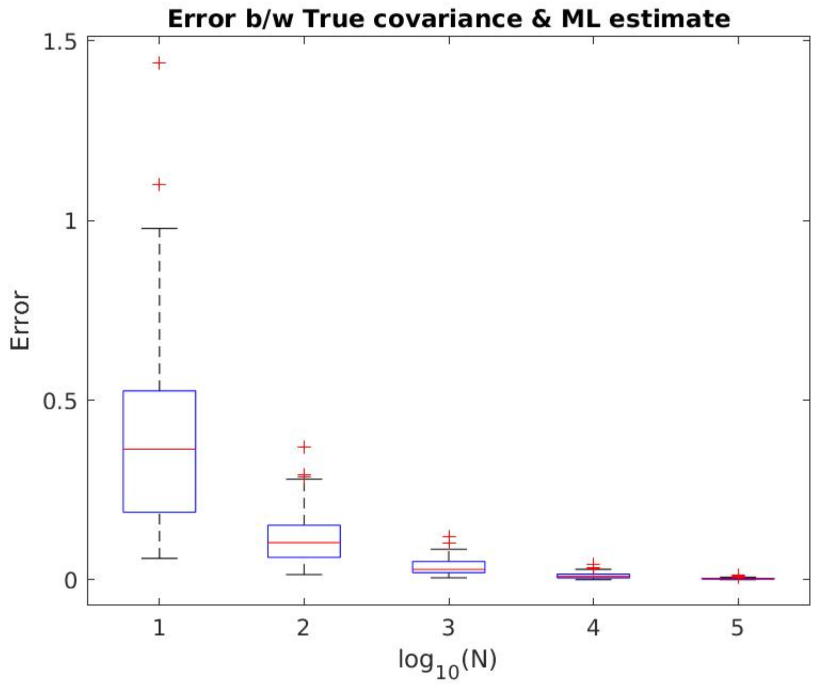

Figure 2: Box plot of the error between the true covariance C and the ML estimate $\hat{C}_N$

(d)

From PCA, we know that principal mode of variation of data lies in the direction of the eigenvector corresponding to highest eigenvalue obtained on Eigen decomposition of the ML Estimate of covariance matrix (here, `cov_mle`).
Again, using `eig` function we obtain the diagonal matrix `lambda` whose diagonal elements are the eigenvalues.
As we only have two eigenvalues, a simple `if-else` can be used to determine the principal mode of variation.
As we use sample mean (here, `mean_mle`) to get the line and plot it using `line`.

```
[Q, lambda] = eig(cov_mle);
if lambda(1, 1) > lambda(2, 2)
    ep = D(1,1)*Q(:, 1);
else
    ep = D(2,2)*Q(:, 2);
end
line([0 ep(1)] + mean_mle(1), [0 ep(2)] + mean_mle(2));
```

Following are the plots for N = 10, 10^2, 10^3, 10^4, 10^5.

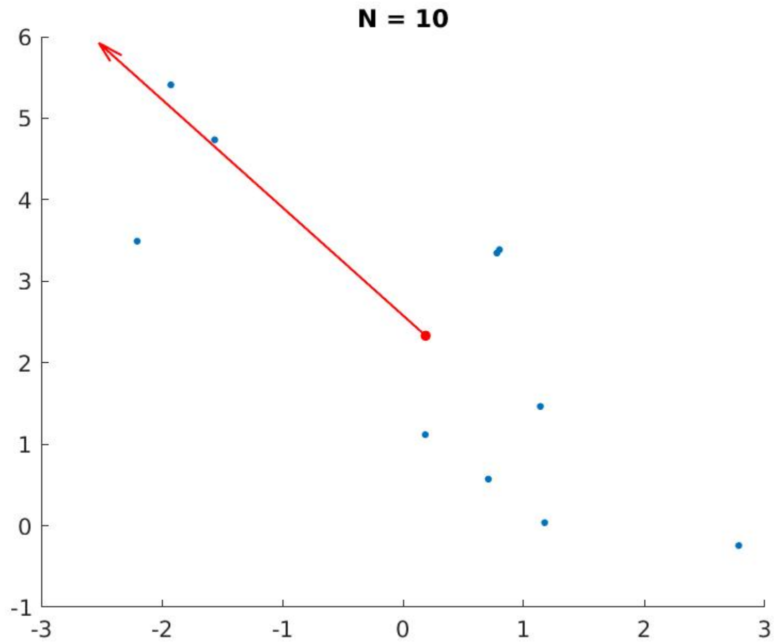

Figure 3: Scatter Plot for $N = 10$ with line showing principal mode of variation

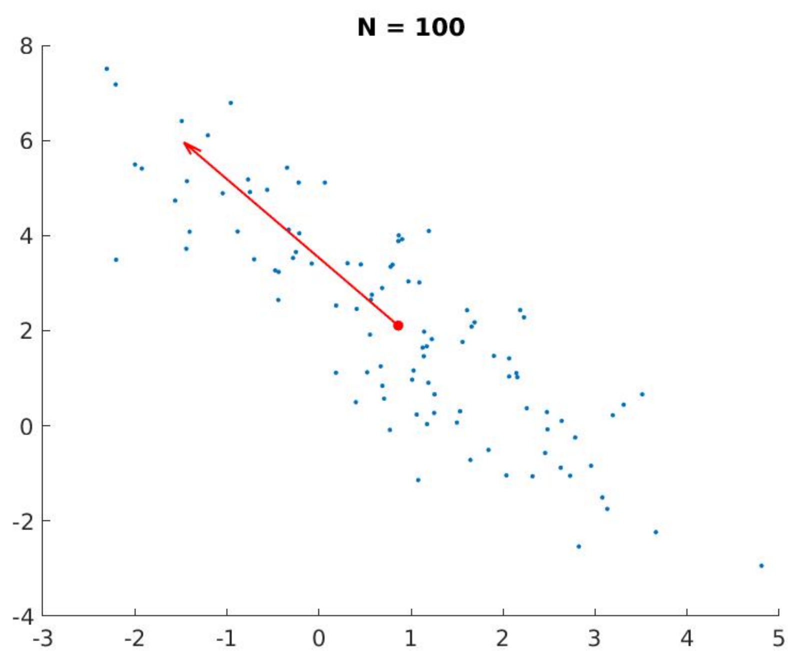

Figure 4: Scatter Plot for $N = 10^2$ with line showing principal mode of variation

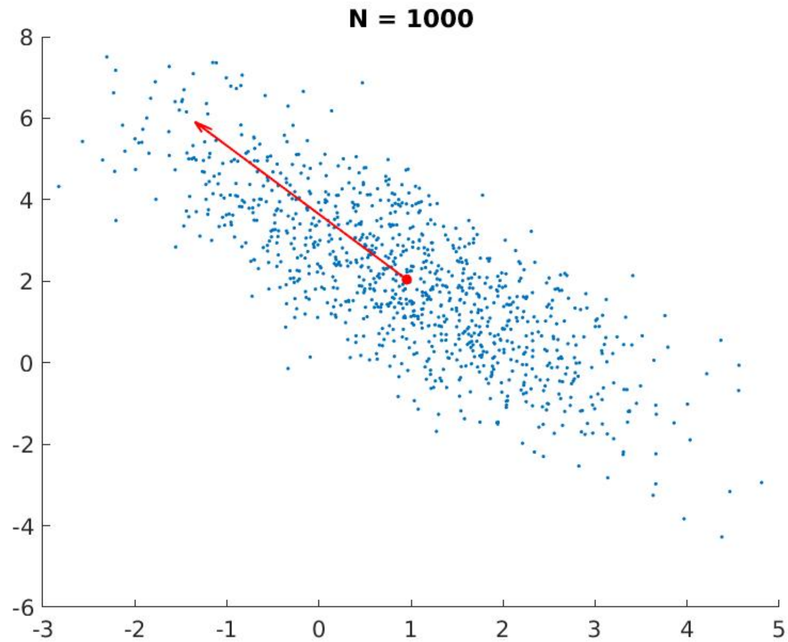

Figure 5: Scatter Plot for $N = 10^3$ with line showing principal mode of variation

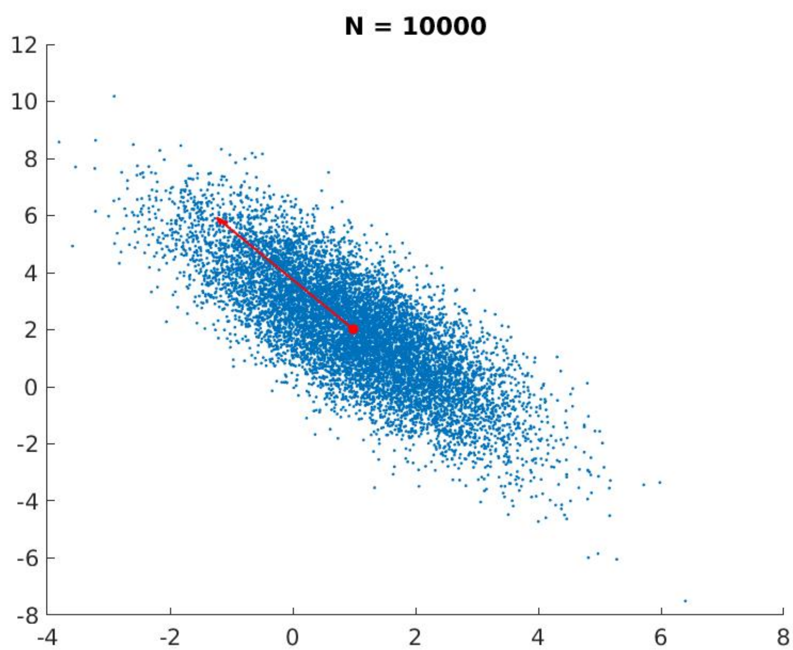

Figure 6: Scatter Plot for $N = 10^4$ with line showing principal mode of variation

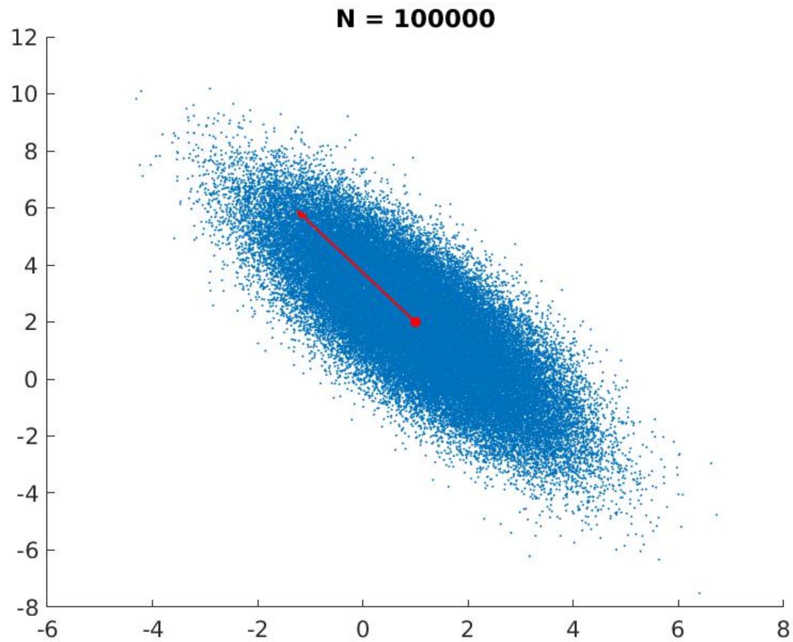

Figure 7: Scatter Plot for $N = 10^5$ with line showing principal mode of variation