“Optimization”

Shervin Halat

98131018

Homework #3

(Implementation Problems)

1.

a.

According to the question, we are to minimize the following expression:

$$\text{minimize} \quad \frac{1}{N}\sum_{i=1}^{N}(d_i - d(x_i, y_i))^2$$

With having 'd' as the following:

$$d(x,y) = \left((x-y)^T P(x-y)\right)^{1/2}$$

Where P is a symmetric PSD matrix.

The objective function can be rewritten in the following form:

$$\text{minimize} \quad \frac{1}{N}\sum_{i=1}^{N}(d_i^2 - 2d_i d(x_i, y_i) + d(x_i, y_i)^2)$$

The objective function is the arithmetic mean on 'I' index where for each index we have the following expression:

$$d_i^2 - 2d_i\left((x_i - y_i)^T P(x_i - y_i)\right)^{1/2} + (x_i - y_i)^T P(x_i - y_i)$$

The first and the last (third) terms are obviously convex in 'P' since the first term is constant and third term is linear in P. The second (middle) term is also convex according to convexity of composition functions theorem since the outer function is a non-decreasing convex function (negation of concave function of square root) composed with another convex linear function; therefore, the whole term is convex and lastly, the whole expression and sum on it will be convex.

Therefore, we have reached a convex optimization problem in 'P' with the following objective function:

$$\text{minimize} \quad \frac{1}{N}\sum_{i=1}^{N}(d_i^2 - 2d_i d(x_i, y_i) + d(x_i, y_i)^2)$$

Subject to $\quad P \in \mathbf{S}_+^n$

b.

In this part, we are to solve the obtained optimization problem above with the following randomly generated mean and covariance matrices:

```
Dataset 1:
   mean: [8.48  2.646]
   covariance:
[[ 7.567  5.387]
 [ 5.387 10.206]]


Dataset 2:
   mean: [8.38  0.649]
   covariance:
[[2.605 3.338]
 [3.338 4.297]]
```

According to the question, two datasets containing 150 two-dimensional datapoints were generated by the mentioned mean and covariance matrices.

For the mentioned parameters, the following error along with 'P' matrix were obtained:

```
P matrix:
[[1. 0.]
 [0. 1.]]

latest error:
0.00000014
```

c.

To evaluate the relation between ‘P’ and covariance matrices of Σ_1 and Σ_2, a coefficient with the values of 0.1, 10 and 100 was defined (namely, covariance coefficient) in order to multiply these values one at a time to simultaneously decrease or increase the values of covariance entries. This way, the covariance matrices will be remained symmetric and PSD. According to the results obtained, as the covariance matrices’ entries increased in value (covariance coefficient increased), the matrix ‘P’ got farther from 2-D identity matrix in addition to increased error and vice versa.

2-D identity matrix: $$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The results are shown below for each of the mentioned ‘covariance coefficient’ values:

“Covariance coefficient of 0.1”

```
P matrix:
[[ 1. -0.]
 [-0.  1.]]

latest error:
0.00000006

covariance coefficient of datasets:
0.1
```

"Covariance coefficient of 10"

```
P matrix:
[[0.892 0.205]
 [0.205 0.883]]

latest error:
1.09888963

covariance coefficient of datasets:
10.0
```

"Covariance coefficient of 100"

```
P matrix:
[[0.418 0.191]
 [0.191 0.49 ]]

latest error:
259.98231796

covariance coefficient of datasets:
100.0
```