Elements of Statistical Learning
Solution Sheet 2

1. (a) The scaling factor $-\frac{1}{2}$ is the same for B and $b_{m,\ell}$, so let's omit it from our consideration for now. The operation

$$-2B = (I_n - \mathbf{1}_n\mathbf{1}_n^T/n)\, E\, (I_n - \mathbf{1}_n\mathbf{1}_n^T/n), \tag{1}$$

can be broken down into four parts. The four parts are really three, as two of the parts are similar. They are

$$A_1 = I_n E I_n, A_2 = -\mathbf{1}_n\mathbf{1}_n^T E I_n/n \text{ and } A_3 = \mathbf{1}_n\mathbf{1}_n^T E\mathbf{1}_n\mathbf{1}_n^T/n^2, \tag{2}$$

which immediately simplifies to

$$A_1 = E, A_2 = -\mathbf{1}_n\mathbf{1}_n^T E/n \text{ and } A_3 = \mathbf{1}_n\mathbf{1}_n^T E\mathbf{1}_n\mathbf{1}_n^T/n^2, \tag{3}$$

The first matrix $A_1 = E$ explains the $e_{m,\ell}$ term in the $b_{m,\ell}$ sum.
Let's pick out the mth row of the matrix $-\mathbf{1}_n\mathbf{1}_n^T E/n$. The mth entry in the first $\mathbf{1}_n$ is just 1, so we need to understand $\mathbf{1}_n^T E$. This is a $1 \times n$ matrix, with each entry equal to summing a column of E and dividing by n. This is a vector containing the average of each column of E and selecting the mth row of the combined matrix selects the average of the mth column. Ditto with the other cross-term which is similar ($E\mathbf{1}_n\mathbf{1}_n^T/n$).
The final component A_3 simultaneously works out the column and row average, ending up the the grand average.

(b) To show that $\mathbf{1}_n$ is an eigenvector of B see that

$$B\mathbf{1}_n = -\frac{1}{2}(I_n - \mathbf{1}_n\mathbf{1}_n^T/n)\, E\, (I_n - \mathbf{1}_n\mathbf{1}_n^T/n), \mathbf{1}_n. \tag{4}$$

Just look at the last bit

$$(I_n - \mathbf{1}_n\mathbf{1}_n^T/n)\mathbf{1}_n = \mathbf{1}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{1}_n = \mathbf{1}_n - \mathbf{1}_n\frac{n}{n} = \mathbf{1}_n - \mathbf{1}_n = \mathbf{0}. \tag{5}$$

So $B\mathbf{1}_n = 0\mathbf{1}_n$ hence $\mathbf{1}_n$ is an eigenvector of B with eigenvalue of 0.

2. (a) A simple answer could be:

```
plot(cmdscale(eurodist), type="n")
text(cmdscale(eurodist), labels=dimnames(cmdscale(eurodist))[[1]],
    cex=0.8)
```

but could be made more complex/prettier with proper axis labels, flipping the axes to make it look more like a proper map of Europe.

(b) Similarly, for `UScitiesD` we can do

```
plot(cmdscale(UScitiesD), type="n")
text(cmdscale(UScitiesD),
    labels=dimnames(cmdscale(UScitiesD))[[1]], cex=0.8)
```

3. (a) $B = XX^T = \begin{pmatrix} 17 & 14 & 13 \\ 14 & 40 & 56 \\ 13 & 56 & 82 \end{pmatrix}$. We use $e_{m,\ell} = b_{m,m} + b_{\ell,\ell} - 2b_{m,\ell}$ to get $E = \begin{pmatrix} 0 & 29 & 73 \\ 29 & 0 & 10 \\ 73 & 10 & 0 \end{pmatrix}$. To obtain the mean of X, just take the mean of each column, which gives $\mu = (2\frac{1}{3}, 5\frac{1}{3})$. The centred data matrix is $X_c = \begin{pmatrix} 1\frac{2}{3} & -4\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \\ -1\frac{1}{3} & 3\frac{2}{3} \end{pmatrix}$. Then $B_c = X_c X_c^T = \frac{1}{9}\begin{pmatrix} 194 & -31 & -163 \\ -31 & 5 & 26 \\ -163 & 26 & 137 \end{pmatrix}$.

(b) I did this in R. My code was:

```
Sheet2Q3 <- function(ret.X=FALSE, ret.B=FALSE, ret.Cm=FALSE,
    ret.E=FALSE, ret.Xmean=FALSE, ret.Xcen=FALSE, ret.Bcen=FALSE,
    ret.Ecen=FALSE, ret.Y=FALSE, ret.By=FALSE)
{
X <- matrix( c(4,1,2,6,1,9), nrow=3, byrow=TRUE)

if (ret.X==TRUE)
return(X)

B <- X %*% t(X)

if (ret.B==TRUE)
return(B)

Cm <- (diag(1, 3) - matrix(c(1,1,1), nrow=3)%*%
    matrix(c(1,1,1), ncol=3)/3)

if (ret.Cm==TRUE)
return(Cm)

E <- matrix(0, 3,3)

for(i in 1:3)
 for(j in 1:3)
  E[i, j] <- B[i,i] + B[j,j] - 2*B[i,j]

if (ret.E==TRUE)
return(E)

Xmean <- apply(X, 2, mean)

if (ret.Xmean==TRUE)
return(Xmean)

Xcen <- Cm %*% X
```

```
if (ret.Xcen==TRUE)
return(Xcen)

Bcen <- Xcen %*% t(Xcen)

if (ret.Bcen==TRUE)
return(Bcen)

Ecen <- E # Just to create matrix of correct dim

for(i in 1:3)
 for(j in 1:3)
  Ecen[i, j] <- Bcen[i,i] + Bcen[j,j] - 2*Bcen[i,j]

if (ret.Ecen==TRUE)
return(Ecen)

Bev <- eigen(Bcen)

f1 <- sqrt(Bev$values[1])*Bev$vectors[,1]
f2 <- sqrt(Bev$values[2])*Bev$vectors[,2]

Y <- cbind(f1, f2)

if (ret.Y==TRUE)
return(Y)


By <- Y %*% t(Y)

if (ret.By==TRUE)
return(By)
}
```

The recovered configuration, Y, is:

```
Sheet2Q3(ret.Y=TRUE)
           f1          f2
[1,]  4.64271  0.02821952
[2,] -0.74144 -0.07630379
[3,] -3.90127  0.04808427
```

The inner product matrix ($\times 9$ to make the numbers easier to see) associated with the recovered configuration:

```
> Sheet2Q5(ret.By=TRUE)*9
     [,1] [,2] [,3]
[1,]  194  -31 -163
```

```
[2,]   -31     5   26
[3,] -163    26  137
```

4. You should have looked at

   `https://en.wikipedia.org/wiki/Seriation_(archaeology)`

5. We prove that if d_1, d_2 are metrics then so is $d = d_1 + d_2$, and the result follows by induction.

   (a) Clearly, $d(x, y) = d_1(x, y) + d_2(x, y) \geq 0$ as both d_1, d_2 are, for objects x, y. If $x = y$ then $d_1(x, y) = d_2(x, y) = 0$, which implies $d(x, y) = 0$.

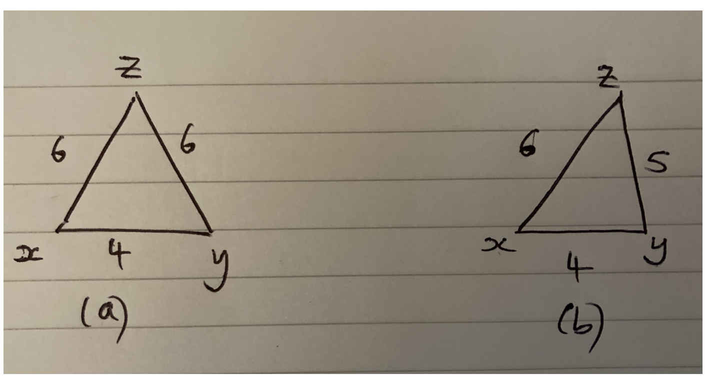
   (b) Also easy that if both d_1, d_2 are symmetric then $d(x, y) = d_1(x, y) + d_2(x, y) = d_1(y, x) + d_2(y, x) = d(y, x)$, so d is symmetric.

   (c) For the triangle inequality we have, for some a, b, c

   $$\begin{aligned} d(a,b) + d(b,c) &= d_1(a,b) + d_2(a,b) + d_1(b,c) + d_2(b,c) && (6)\\ &= d_1(a,b) + d_1(b,c) + d_2(a,b) + d_2(a,b) && (7)\\ &\geq d_1(a,c) + d_2(a,c) && (8)\\ &= d(a,c), && (9) \end{aligned}$$

   as d_1, d_2 are metrics, as required.

6. For the first part, look at the two triangles in the figure

   Triangle (a) is isosceles. It does not matter how you enter the points into the ultrametric inequality, it's always satisfied.

   With triangle (b). Clearly, $4 \leq \max(5, 6)$ and $5 \leq \max(4, 6)$, but $6 > \max(4, 5)$. To satisfy the last inequality, you'd have to increase either the 4 or 5 to 6, and then you're back in the situation as triangle (a). Similarly, if you move that point further than 6, the triangle becomes invalid again for the inequality.

   For the discrete metric.

   (a) Clearly, $\rho(x, y) \geq 0$, since ρ is either zero or one.

   (b) Clearly, $\rho(x, y) = \rho(y, x)$, since it depends on $=$.

(c) By definition $\rho(x, x) = 0$.

(d) By definition $\rho(x, y) = 0$ implies $x = y$.

(e) You only need look at three cases. They are (I) if $x = y = x$, then all discrete metrics between all pairs are zero and the ultrametric inequality is satisfied; (II) $x = y$, but $x \neq z$. Then $\rho(x, y) = 0$ and $\rho(x, z) = \rho(y, z) = 1$ and this satisfies the ultrametric inequality (and corresponds to the isosceles triangle); (III) $x \neq y$ and $y \neq z$ and $x \neq z$. Then $\rho(x, y) = \rho(y, z) = \rho(x, z) = 1$ (and corresponds to the equilateral triangle).

QED.

For those who have time and the interest, this `https://www.colby.edu/math/faculty/Faculty_files/hollydir/Holly01.pdf` was an interesting article on ultrametric spaces. This article is NOT examinable.

[Updated: Feb 19th 2021]