(1, 1pt) Let the matrix be $B = (b_{ij})_{8\times 8}$. Then,

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i\cdot}^2 - d_{\cdot j}^2 + d_{\cdot\cdot}^2),$$

where $d_{i\cdot}^2 = \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2$, $d_{\cdot j}^2 = \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2$, $d_{\cdot\cdot}^2 = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j=1}^{n} d_{ij}^2$.

(2, 1pt) It would be better to have a smaller k such that $P_k = (\sum_{i=1}^{k}\lambda_i)/(\sum_{i=1}^{n}\lambda_i)$ is close to or more than 0.8 (λ_i's are the eigenvalues), for which we would need four to five coordinates. Note that there is no reason to use $\lambda_i > 1$ as selection criterion because $tr(B) \neq n$.

(3, 1pt) The proximity matrix is Euclidean.

(4, 2pts) It shows that it is difficult for the algorithm to distinguish between the three categories using brain scans because the classification powers between any two of the three categories are low (distances between any two of the three points are short).

(5, 2pts) House and Face, because they have the largest *Euclidean distance*, which corresponds to classification power in this case. Note that you should NOT use statistical distance to evaluate the distances between points for the case.

(6, 2pts) Yes, because the proximity matrix is invariant to the rotation of points. Rotation and/or reflection can often be used to facilitate the interpretation of MDS solutions.

(7, 3pts) The dashed axis mainly reflects the *distance* (i.e., classification power) between living (Face and Cat) and non-living (House, Scrambled, Chair, Scissors, Shoes, Bottles) objects. The dotted axis reflects the distance between large and small objects.

(8, 1pt) Under the consideration of the multiple testing encountered here, we would usually allow larger p-values (Note that we only perform the 2nd test when the 1st test is rejected so that the rejection region for the 2nd test in the procedure becomes $RR_1 \cap RR_2$, where RR_1 and RR_2 are the individual rejection regions of the 1st and 2nd tests). Furthermore, there is a significant gap between the p-values of the 2nd and the 3rd tests. Based on these reasons, we should retain the first two canonical pairs.

(9, 2pts) The k-th test statistic is $c\sum_{i=k}^{p} ln(1-\hat{\rho}_i^2)$, where c is a constant. Note that c can be obtained by $\frac{55.41-28.05}{ln(1-0.456)} = -44.94$. The value of the test statistic marked as ?? is $55.41 + (-44.94) \times ln(1-0.884) = 152.22$.

(10, 1pt) Chi-square distribution with degrees of freedom $7 \times 8 = 56$

(11, 2pts) Not consistent. What the researches expected is $\Sigma_{12} = 0$. But, the hypothesis is rejected in the 1st large sample test.

(12, 2pts) All the correlations in R12 are smaller than the first canonical correlation, which is $\sqrt{0.884} = 0.9402$

(13, 3pts) The −FLAVOR1 can be interpreted as the weighted average of CARMEL and AROMA (with weights about 1:2). The TEXTURE1 can be interpreted as the weighted contrast between SURF_RO and LOOSE (with weights about 1.7:1). Note that FLAVOR1 and TEXTURE1 are *negatively* correlated (with correlation −0.9402) as shown in the plot. In summary, the first canonical pair shows that the weighted average of CARMEL and AROMA is positively strongly related to the weighted contrast between SURF_RO and LOOSE.

(14, 2pts) The strength of linearity between FLAVOR_i and TEXTURE_i is measured by their canonical correlation $\hat{\rho}_i$. Because $\hat{\rho}_1^2 \geq \hat{\rho}_2^2 \geq \cdots \geq \hat{\rho}_p^2$, it is expected the scatter plot between FLAVOR1 and TEXTURE1 shows stronger linear relationship than the scatter plot between FLAVOR2 and TEXTURE2. The argument can be applied to any cases of CCA.

(15, 1pt) $R^2 = 45.6\%$

(16, 1pt) equal covariance matrix. Note that the normality assumption is not required in Fisher's approach.

(17, 2pts) Allocate x to 1st group (wild) if $a^T x \geq z12$, where a and $z12$ are given in the question, and to 2nd group (domestic) otherwise.

(18, 2pts) $\frac{|21.161-7.98|}{\sqrt{\frac{13\times 15.986+18\times 11.154}{31}}} = 3.631$, which is the squared root of the Hotelling's T^2 test statistic.

(19, 2pts) No, because we do not know the scales of TIN and other variables. (Note: all variables were measured in millimeters, but their scales might still be different)

(20, 2pts) guilty, because $a^T x_0 = 9.21 < 14.57087$, where $x_0 = (150, 130, 140, 130, 140, 840, 310, 100, 120)$. The turkey is classified as domestic.

(21, 2pts) With this cost ratio, we would tend to classify an accused not-guilty (wild turkey) rather than guilty (domestic turkey). In other word, we would expect to have a cutpoint value smaller than z_{12}. The new value is $z_{12} - ln(100) = 9.966$. We would allocate x to 1st group (wild) if $a^T x \geq 9.966$, and to 2nd group (domestic) otherwise.

(22, 2pts) Two answers are acceptable. The first one is $\frac{4}{33} = \frac{14}{33} \times \frac{2}{14} + \frac{19}{33} \times \frac{2}{19}$, in which the ratio of prior probabilities is $p_1 : p_2 = 14 : 19$. The second one is $\frac{(2/14)+(2/19)}{2}$, in which the ratio of prior probabilities is $p_1 : p_2 = 1 : 1$.

(23, 1pt) Because the accused is classified as guilty in question (20), the change of misclassification for the accused is the estimation of $P(x \in R_2|\pi_1) = P(x \text{ classified as domestic } | \ x \text{ is from wild})$. A straight forward estimate is 2/14.