Consider a multi-class classification problem with 6 classes.

(A) **[Ans]** DDAG requires $6C_2$ pairwise classifiers.

(B) DDAG can not be designed for 6 classes since 6 is not a power of 2.

(C) **[Ans]** DDAG requires 15 pairwise classifiers.

(D) Binary Hierarchical Classification (BHC) is not applicable for this problem, since 6 is not a power of 2.

(E) None of the above.

Consider a multi-class classification problem with 8 classes. Let us compare the following three:

A DDAG with pairwise

B Fully balanced Binary Hierarchical Classifier (BHC)

C Majority voting on pairwise classification.

(A) A, B, and C will require exactly the same number of classifiers.

(B) **[Ans]** A is faster than C (A requires less compute than C) for evaluating/testing a sample.

(C) **[Ans]** B is faster than A and C (B requires less compute than A and C) for evaluating/testing a sample.

(D) **[Ans]** C is better suited for parallel evaluation than A and B

(E) All the above statements are true.

Consider an 8-class classification with Binary Hierarchical classification (BHC).

(A) We prefer Balanced BHC, since balanced BHCs will have the highest accuracy.

(B) If BHC is not balanced, we will have multiple leaves with the same label.

(C) If BHC is not balanced, number of classifiers will increase (compared to the balanced one)

(D) **[Ans]** If BHC is not balanced, average time for classification (amount of compute) will increase (compared to the balanced one).

(E) An unbalanced BHC can be converted to a BHC with no loss in accuracy with some rotate operators (just like a rotate operations in AVT Tree in a typical data structure course)

(F) All the above.

Consider a multi-class classification problem with K classes.

We have now K one vs rest linear classifiers are designed as $\mathbf{w_1} \dots, \mathbf{w_K}$

(A) We prefer "Classify as k if $\mathbf{w_k^T x} \geq 0$". This will have unambiguous and correct classification.

(B) **[Ans]** We prefer "Classify as k if k is $\arg\max_k \mathbf{w}_k^T \mathbf{x}$".

(C) **[Ans]** Finding $\mathbf{w_1} \dots, \mathbf{w_K}$ can be formulated and solved as K independent training problem.

(D) Finding $\mathbf{w_1} \dots, \mathbf{w_K}$ has to be formulated and solved as a single training/optimization problem.

(E) **[Ans]** We used "Classify as k if k is $\arg\max_k \mathbf{w}_k^T \mathbf{x}$" and this resulted in all samples correctly classifying with no ambiguity. If this is the case, all the $\mathbf{w}_i$ (say in a 2D plane) geometrically define lines that intesect at a common point.

Consider a K class multi-class classifier implemented with pair-wise classifier and majority voting.

Accuracy of samples in class ω_i is η_i.

(A) **[Ans]** Final decision is the class that gets majority votes.

(B) Overall accuracy is the sum of accuracies of all the K classes. i.e., $\sum_{i=1}^{K} \eta_i$

(C) Overall accuracy is the average of accuracies of all the K classes. i.e., $\frac{1}{K}\sum_{i=1}^{K} \eta_i$

(D) **[Ans]** Overall accuracy is the weighted average of accuracies of all the K classes, where weights are the prior probabilities of each of the classes i.e., $\frac{1}{K}\sum_{i=1}^{K} P(\omega_i)\eta_i$

(E) Overall accuracy is the weighted average of accuracies of all the K classes, where weights are the inverse of the prior probabilities of each of the classes i.e., $\frac{1}{K}\sum_{i=1}^{K} \frac{1}{P(\omega_i)}\eta_i$