Q1

Since the age is a numeric value between 0 and 120, I want to use regression tree.

Consider the weighted variance reduction:

$$\text{score}_{var} = var(T) - (\frac{N_{T_L}}{N_T} Var(T_L) + \frac{N_{T_R}}{N_T} Var(T_R))$$

Where for any node T, var(T) is the variance of node samples:

$$var(T) = \frac{1}{N_T} \sum (y_i - \overline{y_T})$$

I will enumerate all the possible splitting points for each attribute and choose the split point that has maximum score.

Q2

Yes, I agree. When weak learner is better than random guess is greater than random guess, the accumulation of weak learner will cause overfitting. Therefore, it will lead to a nontrivial increase of classification accuracy.

Q3

The dimension of data is high, so I prefer using SVM, since it's suitable for high dimensional data.

Naïve Bayes and Decision Tree are not suitable for this data. For Naïve Bayes, it assumes each feature is independent to each other, while the assumption may be violated in this data. For Random Forest, since the number of feature is too large, it will cost much time to train the model, also, overfitting may happen.