Question 1.

What is the null hypothesis and the alternative hypothesis in each of the statistical test?

Paired t-test:

Null Hypothesis(H0):True Difference in means is equal to 0.

Alternative hypothesis(H1):True Difference in means not equal to 0.

Anova Test:

Null Hypothesis(H0): Differences in errors for different classifiers are significantly different.

Alternative Hypothesis(H1): Differences in errors for different classifiers are approximately equal.

Wilcox Signed Rank Test:

Null Hypothesis: True location shift is equal to 0.

Alternative hypothesis: True location shift is not equal to 0.

Question 2.

What did you conclude after performing the tests? State which hypothesis seemed favorable given the evidence.

If p-value is greater than 0.05 we don't have sufficient evidence to reject null hypothesis otherwise null hypothesis is rejected.

Paired t-test:

For the paired t-test, obtained p-value of 0.3434, therefore we don't have sufficient evidence to reject null hypothesis. True Difference in means is equal to 0.

Anova test:

Since, p-value is (0.0285) less than 0.05, therefore, we can reject the null hypothesis. Differences in errors for the four classifiers are approximately equal.

Wilcox Signed Rank Test:

For the wilcox signed rank test, p value is 0.1362 which is greater than 0.05. Hence we can't reject the null hypothesis. Hence True location shift is equal to 0.

Question 3.

Briefly explain why signed rank test is more useful than ANOVA test while comparing two classifier algorithms on multiple data sets.

1. The main difference between these two tests are that ANOVA is a parametric test while Wilcox Signed rank test is a non-parametric test. i.e the Wilcox Signed rank test doesn't assume your error dataset to follow some distribution.
2. Also, the Wilcox Signed rank Test is advantageous in situations where you want to compare other stats such as training times, no of parameters, etc.
3. If we compare the training time of two algorithms of two datasets in which one of the dataset is significantly larger than the other, then using ANOVA would skew our measurements towards the larger dataset as compared to Wilcox test.
4. The Wilcoxon is vastly superior to the ANOVA in most cases, especially with non-normal data (read anything by Cliff Blair or Shlomo Sawiliowsky for references).
5. The ANOVA is somewhat robust to non-normality, but not very.
6. Additionally the Wilcoxon (as with most rank based statistical tests) are more powerful and increase in power with non-normal data. However, some fields tend to have a bias against distribution free statistical tests. This bias is founded in the tradition and NOT the math.
7. Both statistical tests are the same in R, SPSS, SAS, Minitab, etc.
8. Another disadvantage of anova is if the dataset is divided into multiple datasets, Anova gets biased towards bigger datasets in terms of its results as compared to smaller counterparts.