Machine Learning - Solutions for Evaluation of Hypotheses

7.1 Solution:
We are given:
• Test set size (n) = 45
• Error rate observed = 6.67% = 0.0667
Estimated error (p) = 3/45 = 0.0667
Standard error (SE) = sqrt(p(1-p)/n) = sqrt(0.0667 * 0.9333 / 45) ≈ 0.0372
Using z = 1.96 for 95% confidence interval:
CI = p ± z * SE = 0.0667 ± 1.96 * 0.0372 ≈ [-0.006, 0.140]
Since error rate cannot be negative, final CI = [0, 0.14] or [0%, 14%].

7.2 Solution:
h1 error = 6.67%, h2 error = 8.89%, h3 error = 13.3%.
We compare differences:
• Difference h2 - h1 = 2.22%.
• Difference h3 - h1 = 6.63%.
Larger observed difference (h3 vs h1) gives higher confidence that h3 is worse.
Thus, we are more confident that h3 performs worse than h1 compared to h2 vs h1.

7.3 Solution:
We apply a paired t-test using the 10-fold CV errors:
Differences (Favourite - Decision Tree): [-0.41, 0.04, -0.99, 0.35, 0.14, -0.78, -0.46, 0.56, 0.00, 0.03].
Mean difference ≈ -0.152%.
Standard deviation ≈ 0.50%.
t = mean / (sd/sqrt(10)) ≈ -0.96.
For df = 9, critical t at 95% = 2.26.
Since |t| < 2.26, difference is not significant.
Conclusion: Cannot claim with 95% confidence that favourite algorithm outperforms decision tree.
Confidence is low.

7.4 Solution:
We compute ROC coordinates:
For h1: TPR = 29/31 = 0.935, FPR = 1/14 = 0.071.
For h2: TPR = 29/30 = 0.967, FPR = 3/15 = 0.200.
For h3: TPR = 27/30 = 0.900, FPR = 3/15 = 0.200.

Euclidean distance to perfect classifier (0,1):
• h1: sqrt((0.071-0)^2 + (0.935-1)^2) ≈ 0.103.
• h2: sqrt((0.200-0)^2 + (0.967-1)^2) ≈ 0.203.
• h3: sqrt((0.200-0)^2 + (0.900-1)^2) ≈ 0.224.
=> Best is h1 under equal costs.

If false positives cost 4x more:
Weighted distance metric: D = sqrt((4*FPR)^2 + (1-TPR)^2).
• h1: sqrt((0.284)^2 + (0.065)^2) ≈ 0.291.
• h2: sqrt((0.800)^2 + (0.033)^2) ≈ 0.801.
• h3: sqrt((0.800)^2 + (0.100)^2) ≈ 0.806.
=> Best is still h1.
Final Answer: h1 is the best classifier in both cases.