Machine Learning

Exercise Sheet on Evaluation of Hypotheses

Q1.

Given:

- Test set size = **45 instances**
- Observed error rate = **6.67% = 0.0667**
- Confidence level = **95%**

Confidence Interval = $\widehat{}\{p\} \pm z \cdot \sqrt{\{\backslash frac\{\widehat{}\{p\}(1 - \widehat{}\{p\})\}\{n\}\}}$

Standard Error (SE) = √ ((450.0667 * (1 − 0.0667)) / 45)

= √ ((450.0667 * 0.9333))

= √ (450.06227)

= √ (0.0013848) ≈ 0.03721

Now calculate the margin of error: = Margin = Z * SE = 1.96 * 0.03721≈0.07293

Lower Bound=p^ − Margin = 0.0667 − 0.07293 = − 0.0062
(clip to 0, as error can't be negative)

Upper Bound = p^ + Margin = 0.0667 + 0.07293 = 0.1396

95% Confidence Interval for the true error is [0.00,0.1396] = [0%,13.96%].

Q2.

Given Data

- Number of test instances, n = 45
- Observed error rate for h1, p_1 = 6.67% = 0.0667
- Observed error rate for h2, p_2 = 8.89% = 0.0889

- Observed error rate for h3, p_3 = 13.3% = 0.133

  SE = √ (((p1(1 – p1) / n) + p2(1 – p2) / n)

  SE $_{1,2}$ = √ ((0.0667 * 0.9333 / 45) +(0.0889 * 0.9111 / 45)

  = √ (0.06227 / 45 + 0.081045) = √ (0.0013848 + 0.0018)

  = √ 0.0031848 ≈ 0.0564

  Z-Score = (p2−p1)/SE = (0.0889−0.0667) / 0.0564 ≈0.394

  For comparison between h1 and h3:

  SE $_{1,3}$ = √ ((0.0667 * 0.9333 / 45) + (0.133 * 0.867 / 45))

  = √ (0.0013848 + 0.002564) = √ 0.0039488≈0.06280.133

  Again, 1.054 < 1.645 → **not statistically significant** at 95%, but **higher than for h2**.

Q3.

Let the differences be:

d = [−0.63, −1.35, −0.11, −0.56,0.12,0.36, −0.85,1.99, −0.67,0.08]

Mean of differences:

D = ∑d_i / n = −1.6210

Standard deviation (s) of differences:

First calculate:

∑ (d_i− D)^2 = 6.94436

Then, s = √ (6.94436 / (10−1)) = √ 0.7716 ≈ 0.8786

Perform Paired t-test

T = D (s/√n = − 0.162 / (0.8786/ √10 ≈ −0.162 / 0.2778≈−0.583

Degrees of freedom = 10 − 1 = 9

|t|=0.583<1.833 => We **fail to reject the null hypothesis**.

Confidence level is low — You cannot confidently claim that one algorithm is significantly better than the other in this domain based on the provided cross-validation results.

Q4.

Now, let's calculate the True Positive Rate (TPR) and False Positive Rate (FPR) for each classifier:

Classifier 1:

$TPR = TP/(TP+FN) = 29/31 = 0.9355$

$FPR = FP/(FP+TN) = 1/14 = 0.0714$

Classifier 2:

$TPR = TP/(TP+FN) = 29/30 = 0.9667$

$FPR = FP/(FP+TN) = 3/15 = 0.2000$

Classifier 3:

$TPR = TP/(TP+FN) = 27/30 = 0.9000$

$FPR = FP/(FP+TN) = 3/15 = 0.2000$

The perfect classifier in an ROC plot would be at point (0,1), meaning $FPR = 0$ and $TPR = 1$.

For equal costs of false positives and false negatives:

Let's calculate the Euclidean distance from each classifier to the perfect point (0,1):

Classifier 1: Distance $= \sqrt{[(0.0714-0)^2 + (0.9355-1)^2]} = \sqrt{[0.0051 + 0.0042]}$

$= \sqrt{0.0093} = 0.0964$

Classifier 2: Distance $= \sqrt{[(0.2000-0)^2 + (0.9667-1)^2]} = \sqrt{[0.0400 + 0.0011]}$

$= \sqrt{0.0411} = 0.2027$

Classifier 3: Distance $= \sqrt{[(0.2000-0)^2 + (0.9000-1)^2]} = \sqrt{[0.0400 + 0.0100]}$

$= \sqrt{0.0500} = 0.2236$

Since Classifier 1 has the smallest distance (0.0964), it's the best classifier when costs are equal.

For false positives costing 4 times as much as false negatives:

Classifier 1: Weighted Distance $= \sqrt{[(4 * (0.0714-0))^2 + (0.9355-1)^2]}$

$= \sqrt{[4^2 \times 0.0051 + 0.0042]} = \sqrt{[0.0816 + 0.0042]} = \sqrt{0.0858} = 0.2929$

Classifier 2: Weighted Distance = $\sqrt{[(4 * (0.2000-0))^2 + (0.9667-1)^2]}$

$= \sqrt{[4^2 \times 0.0400 + 0.0011]} = \sqrt{[0.6400 + 0.0011]} = \sqrt{0.6411} = 0.8007$

Classifier 3: Weighted Distance = $\sqrt{[(4 * (0.2000-0))^2 + (0.9000-1)^2]}$

$= \sqrt{[4^2 \times 0.0400 + 0.0100]} = \sqrt{[0.6400 + 0.0100]} = \sqrt{0.6500} = 0.8062$

Once again, Classifier 1 has the smallest weighted distance (0.2929), making it the best classifier even when false positives cost 4 times as much as false negatives.

For both scenarios (equal costs and when false positives cost 4 times more than false negatives), Classifier 1 is the best choice based on the Euclidean distance from the perfect classifier in the ROC plot.