Question no 4: Using the Euclidean distance on an ROC plot from the "perfect classifier" as the metric, choose the best classifier

• assuming equal costs for false positives and false negatives

• assuming that false positives cost 4 times as much as false negatives

h1 Confusion Matrix

Actual	Positive	Negative	Marginal Sum
Predicted Positive	29	1	30
Negative	2	13	15
Marginal Sum	31	14	45

h2 Confusion Matrix

Actual	Positive	Negative	Marginal Sum
Predicted Positive	29	3	32
Negative	1	12	13
Marginal Sum	30	15	45

h3 Confusion Matrix

Actual	Positive	Negative	Marginal Sum
Predicted Positive	27	3	30
Negative	3	12	15
Marginal Sum	30	15	45

To choose the best classifier, we will calculate the True Positive Rate (TPR), False Positive Rate (FPR), and then the Euclidean distance from the "perfect classifier" (0,1) on an ROC plot.

Formulas

True Positive Rate (TPR) / Sensitivity: TP / (TP + FN)

False Positive Rate (FPR): FP / (FP + TN)
Euclidean Distance (d): sqrt((FPR - 0)^2 + (TPR - 1)^2)

Hypothesis h1

- TP = 29, FN = 2
- FP = 1, TN = 13

- TPR = 29 / (29 + 2)
  = 29 / 31
  = 0.9355

- FPR = 1 / (1 + 13)
  = 1 / 14
  = 0.0714

- Euclidean Distance = sqrt((0.0714 - 0)^2 + (0.9355 - 1)^2)
  = sqrt(0.0051 + 0.0042)
  = sqrt(0.0093)
  = 0.0964

Hypothesis h2

- TP = 29, FN = 1
- FP = 3, TN = 12

- TPR = 29 / (29 + 1)
  = 29 / 30
  = 0.9667

- FPR = 3 / (3 + 12)
  = 3 / 15
  = 0.2000

- Euclidean Distance = sqrt((0.2000 - 0)^2 + (0.9667 - 1)^2)
  = sqrt(0.0400 + 0.0011)
  = sqrt(0.0411)
  = 0.2027

Hypothesis h3

- TP = 27, FN = 3
- FP = 3, TN = 12

- TPR = 27 / (27 + 3)
  = 27 / 30
  = 0.9000

- FPR = 3 / (3 + 12)
  = 3 / 15

= 0.2000

- Euclidean Distance = sqrt((0.2000 - 0)^2 + (0.9000 - 1)^2)
  = sqrt(0.0400 + 0.0100)
  = sqrt(0.0500)
  = 0.2236

Summary of Euclidean Distances (Equal Costs)

Hypothesis	TPR	FPR	Euclidean Distance from Perfect Classifier
h1	0.9355	0.0714	0.0964
h2	0.9667	0.2000	0.2027
h3	0.9000	0.2000	0.2236

Based on the calculations, **h1** has the smallest Euclidean distance (0.0964). Therefore, **h1 is the best classifier assuming equal costs for false positives and false negatives.**

b) Assuming false positives cost 4 times as much as false negatives:

In this case, we modify the distance formula for false positives more heavily:
Distance= sqrt($(4×FPR)^2+(1−TPR)^2$)

h1 : Distance = sqrt($(4×0.071)^2 +(1−0.935)^2$)
=sqrt(0.284^2 +0.065^2)
= sqrt(0.080656+0.004225)
= sqrt(0.084881)
= 0.291

h2 : Distance = sqrt($(4×0.2)^2 +(1−0.906)^2$)
=sqrt(0.8^2 +0.094^2)
= sqrt(0.64+0.008836)
= sqrt(0.648836)
= 0.805

h3 : Distance = sqrt($(4×0.2)^2 +(1−0.9)^2$)
=sqrt(0.8^2 +0.1^2)
= sqrt(0.64+0.01)
= sqrt(0.65)
= 0.806

Even with the higher penalty on false positives, **h1** is still the best classifier because its FPR is significantly lower than h2 and h3.
Therefore, **h1 is also the best classifier assuming that false positives cost 4 times as much as false negatives**.