4. Choose the best classifier using Euclidean distance on an ROC plot

The "perfect classifier" on an ROC plot is at the top-left corner (False Positive Rate, FPR = 0; True Positive Rate, TPR = 1). The Euclidean distance D from the perfect classifier (0,1) is:

D = **Root(**(FPR − 0)^2 + (TPR − 1)^2 **)**

= **Root(**FPR^2 + (1 − TPR)^2**)**

The classifier with the minimum distance D is the best.

From the confusion matrices:

Hypothesis	True Positive (TP)	False Negative (FN)	False Positive (FP)	True Negative (TN)	Actual Positive (P)	Actual Negative (N)
h1	29	2	1	13	31	14
h2	29	1	3	12	30	15
h3	27	3	3	12	30	15

Metrics Calculation:

- **True Positive Rate (TPR):** Recall = PTP = TP + FNTP
- **False Positive Rate (FPR):** NFP = FP + TNFP

Case A: Equal Costs for False Positives (FP) and False Negatives (FN)

This case uses the standard Euclidean distance for the perfect classifier (0,1) on the ROC plot.

1. Calculate TPR and FPR for h1:

- P = 31, N = 14
- TPR(1) = 29 / 31 = 0.9355
- FPR(1) = 1 / 14 = 0.0714
- D1 = **Root(**0.07142+(1−0.9355)^2**)**

= **Root(**0.0051+0.0042**)**

= **Root(**0.0093**)**

= 0.0964

2. Calculate TPR and FPR for h2:

- P = 30, N = 15
- TPR(2) = 29 / 30 = 0.9667
- FPR(2) = 3 / 15 = 0.2000
- D2 = **Root(**0.20002 + (1 − 0.9667)^2 **)**

= **Root(**0.0400 + 0.0011 **)**

= **Root(**0.0411**)**

= 0.2027

3. Calculate TPR and FPR for h3:

- P = 30, N = 15
- TPR(3) = 27 / 30 = 0.9000
- FPR(3) = 3 / 15 = 0.2000
- D3 = **Root(**0.20002 + (1 − 0.9000)^2**)**

= **Root(**0.0400 + 0.0100)

= **Root(**0.0500)

= 0.2236

Best Classifier (Equal Costs): h1 has the minimum distance (D1 = 0.0964).

Case B: False Positives Cost 4 times as much as False Negatives

When costs are unequal, the cost-sensitive distance from the perfect classifier is used. If CFP is the cost of a false positive and CFN is the cost of a false negative, the weighted distance Dw is:

Dw = **Root(** (FPR×CFP / Cost Norm)^2 + ((1−TPR)×CFN / Cost Norm)^2**)**

Here, CFP = 4C(FN). Let C(FN) = 1, then C(FP) = 4.

The distance from the perfect classifier is given by the modified distance metric which weighs the axes:

Dw = **Root(**(4 × FPR)^2 + (1 × (1 − TPR))^2**)**

1. Weighted Distance for h1:

- FPR1 = 0.0714, 1 − TPR1 = 0.0645
- Dw,1 = **Root(**(4×0.0714)2+0.06452)

  = **Root(**0.28562+0.06452)

  = **Root(**0.0816+0.0042)

  = **Root(**0.0858)

  = 0.2929

2. Weighted Distance for h2:

- FPR2 = 0.2000, 1−TPR2 = 0.0333
- Dw,2 = **Root(**(4×0.2000)2+0.03332)

  = **Root(**0.80002+0.03332)

  = **Root(**0.6400+0.0011)

  = **Root(**0.6411)

  = 0.8007

3. Weighted Distance for h3:

- FPR3 = 0.2000, 1−TPR3 = 0.1000
- Dw,3 = **Root(**(4×0.2000)2+0.10002)

  = **Root(**0.80002+0.10002)

  = **Root(**0.6400+0.0100)

  = **Root(**0.6500)

  = 0.8062

Best Classifier (FP cost 4x FN): h1 has the minimum weighted distance (Dw,1 = 0.2929).

For **Equal Costs (Case A)** the best classifier is **h1**.

For **FP Cost 4x FN (Case B)** best classifier is **h1**.