Answer **A**:

I. According to the Cost matrix the best **classifier is A**

Explanation:

In Cost matrix the high value is 80, which comes when an actual NO predicted as YES. The chances of getting that error higher in **Classifier B (30).**

Negative value in a Cost matrix interrupted as benefits and positive values are benefits. Hence in the diagonal elements are the correct prediction, which gives the real benefits.

II. **Cohen's kappa** is defined as:

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

Kappa Value of Classifier A

	A	B	Total
A	50	10	**60**
B	5	100	**105**
Total	**55**	**110**	**165**

Number of observed agreements: 150 (90.91% of the observations)
Number of agreements expected by chance: 90.0 (54.55% of the observations)

Kappa= 0.800
SE of kappa = 0.049
95% confidence interval: From 0.704 to 0.896

"One way to interpret kappa is with this scale (1):

Kappa < 0: No agreement
Kappa between 0.00 and 0.20: Slight agreement
Kappa between 0.21 and 0.40: Fair agreement
Kappa between 0.41 and 0.60: Moderate agreement
Kappa between 0.61 and 0.80: Substantial agreement
Kappa between 0.81 and 1.00: Almost perfect agreement."

Kappa Value of Classifier B

	A	B	Total
A	30	30	**60**
B	20	85	**105**
Total	**50**	**115**	**165**

Number of observed agreements: 115 (69.70% of the observations)
Number of agreements expected by chance: 91.4 (55.37% of the observations)

Kappa= 0.321
SE of kappa = 0.077
95% confidence interval: From 0.171 to 0.471

"One way to interpret kappa is with this scale (1):

Kappa < 0: No agreement
Kappa between 0.00 and 0.20: Slight agreement
Kappa between 0.21 and 0.40: Fair agreement
Kappa between 0.41 and 0.60: Moderate agreement
Kappa between 0.61 and 0.80: Substantial agreement
Kappa between 0.81 and 1.00: Almost perfect agreement."

According to Kappa Value Classifier A is the best classifier.

III. Classifier A

Matrix for predicting cancer

		PREDICTED CLASS	
		Class = Yes	Class = No
ACTUAL CLASS	Class = Yes	a (TP)	b (FN)
	Class = No	c (FP)	d (TN)

a = TP (true positive)
b = FN (false negative)
c = FP (false positive)
d = TN (true negative)

Precision = a/ (a+c)

= 50 / (50+5)

=0.9090

Recall = (a)/(a+b)

=50 / (60)

=0.8333

F- Measure = 2a / (2a+b+c)

=100 / (100+10+5)

= 0.869

Classifier B

Precision = a/ (a+c)

= 30 / (30+20)

=0.60

Recall = (a)/(a+b)

=30 / (60)

=0.50

F- Measure = 2a / (2a+b+c)

=60 / (60+30+20)

= 0.545

IV. Accuracy

Classifier A

ACC = (a + d) / (P + N)

Accuracy= 0.90

Classifier B

Accuracy = 0.69

Answer **B:**

I. Size of instances for the below attached data is 4

i

Sr.	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Marked in **RED is called instance.**

Marked **BLUE** is called Feature.

II. **Size of hypothesis size is 2**

III. **S algorithm**

h0 = {sunny,warm,normal,stron,warm,same}

h1 = {sunny,warm,high,strong,warm,same} - > {?,warm,?,strong,,warm,same}

h2 = {?,warm,?,Strong,?,?}

h3 = {?,?,?,Strong,?,?} - > Final Hypothesis

IV. When adding an attribute (row) Waterflow, which is having 3 different values that will affect hypothesis spaces only. Number of instances will be changed once the column size changed.

Size of hypothesis will be 3