

6. a. A relational database has the following columns for a student table:

Student ID (primary key)	Student Name	Student's Department	Advisor ID	Advisor Name
-----------------------------	--------------	----------------------	------------	--------------

Explain why this table is not in third normal form

Transitive dependency exists: advisor name depends on advisor ID, not student ID

b. Explain, or draw a new table or tables for how you would normalize this database to third normal form:

Remove transitive dependencies: Split table so that all non-primary key columns are dependent only on the primary key of its table

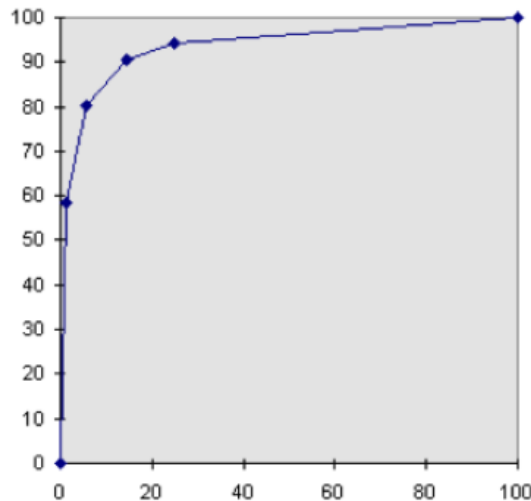
Student ID (primary key)	Student Name	Student's Department	Advisor ID
-----------------------------	--------------	----------------------	------------

Advisor ID	Advisor Name
------------	--------------

7. eQTLs are genomic loci that explain variation in _____

mRNA expression levels

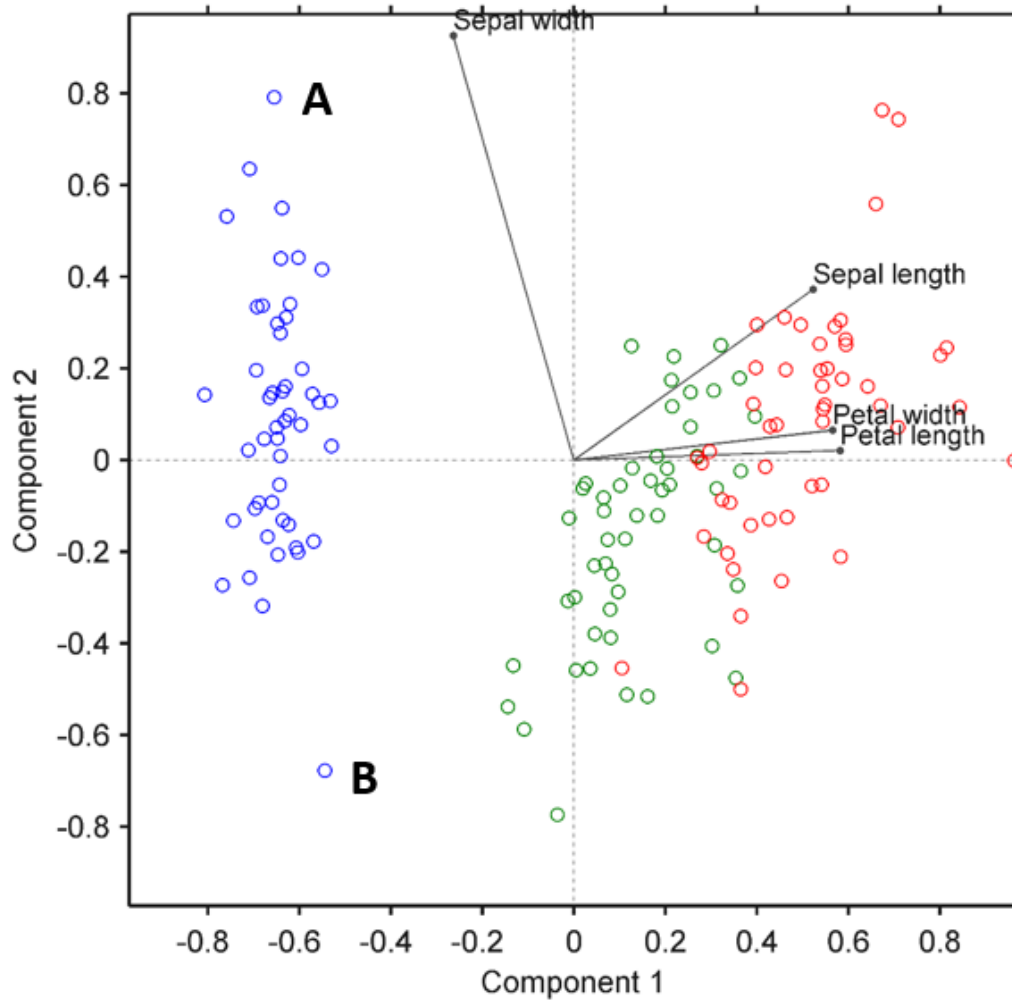
8. In the following ROC (Receiver Operating Curve), the area under the curve indicates that this classifier that performs better than random guessing. Label the axes and also define each axis in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)



y-axis: $TPR = TP / (TP + FN)$

x-axis: $FPR = FP / (TN + FP)$ or $1 - TN / (TN + FP)$

9.



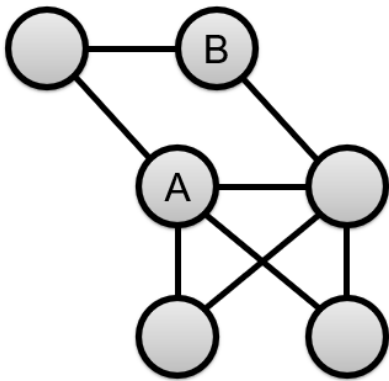
For the above PCA biplot, assume that the first 2 components account for 95% of the variance in the original dataset. Answer True or False:

Petal width and petal length are strongly correlated **True**

Sepal width and petal width are strongly correlated **False**

Petal length strongly projects onto principal component 2 **False**

Points A and B differ mostly in sepal width **True**



10. For the above network, what are the following values:

The degree of node A?

4

The clustering coefficient of node A?

$2/6$ or $1/3$

The clustering coefficient of node B?

0

The shortest path length from A to B?

2