2018AIML600

**Ans 1.**

Principal components analysis (PCA) is one of a family of techniques for taking

high-dimensional data, and using the dependencies between the variables to represent

it in a more tractable, lower-dimensional form, with preserving information as much as possible.

PCA is one of the simplest and most robust ways of doing such dimensionality.

PCA enables us to reduce as many dimensions from dataset as possible with preserving their variance as much possible.

In PCA, we can reduce like 7 dimensions vectors to 1 or 2 dimensions as per requirement.

On the other hand, aggregation is groups of dimensions are combined. Like weight and heights of a person can be combined as body mass indexes.

It requires some domain expertise to combine dimensions in that way and it cannot be applied to any two or more dimensions but to specific only.

**Ans 2:**

Eigenvalues of a covariance matrix, indicates the magnitude of principal component. Larger eigenvalues means more variance while smaller values mean lesser variance among the values. So in turn, small eigenvalues means actual data are not dispersed in elliptical manner but to a circular manner. Correlation between feature and target is not good enough to determine.

**Ans 3.**

Feature Subset Selection, is the process of selecting a subset of relevant features for use in model construction. It is basically depends upon the score, which is measured as a distance between the desired output and observed output. Subset selection algorithms can be broken up into Wrappers, Filters and Embedded

Principal components analysis (PCA), on the other hand, is one of a family of techniques for taking

high-dimensional data, and using the dependencies between the variables to represent

it in a more tractable, lower-dimensional form, with preserving information as much as possible.

PCA is one of the simplest and most robust ways of doing such dimensionality.

PCA enables us to reduce as many dimensions from dataset as possible with preserving their variance as much possible.

Ans 4

(a)

PC2



PC2

PC1

PC1



Angle between both PC1 and PC2 in both figures = 900

(b) We can see that. In figure 1, PC1 can separate the classes more accurately, while in figure 2, PC2 separate the both classes more accurately, it depends upon the data distributions. PCA is not suitable for separating two classes, while LDA is good for that particular purpose. PCA mainly depends upon the covariance of the projections on component vectors irrespective of the class of which data belongs. So we should not rely on PCA for class distinguish purpose.