SMAI-M20-06: Data, Distances and Learning

C. V. Jawahar

IIIT Hyderabad

August 21, 2020

Review Question - I (one, none or more correct)

If

$$Ax = \lambda x$$

then What are the eigen values and eigen vectors of A^2
i.e., Find:

$$A^2 x = ?$$

- (a) λ, x
- (b) λ^2, x
- (c) λ, $2x$
- (d) λ^2, $2x$
- (e) none of the above

Review Question - II (one, none or more correct)

Consider

$$\mathbf{x}_i' = \mathbf{W}\mathbf{x}_i$$

$\mathbf{W}$ is constructed as below:

- We start with a $d \times d$ identity matrix
- We randomly permute (rearrange) the <u>columns</u>
- We remove half of the <u>rows</u> and create a $\frac{d}{2} \times d$ matrix $\mathbf{W}$

The process of creation of new representation is:

- (a) Feature subset selection; A random subset of the original features will be in $\mathbf{x}'$
- (b) Feature extraction; New features are linear combination of old ones, and not really a subset.
- (c) Dimensionality Reduction; New representation has smaller dimension than the original one.
- (d) This can not be done since these operations are illegal (or mathematically not defined).

Review Question - III (one, none or more correct)

Consider three sets

$$A = \{1, 3, 4, 5, 6, 7, 8\}$$

$$B = \{2, 4, 6, 8\}$$

$$C = \{1, 2, 3, 4, 5\}$$

We know Jacard index ($J = \frac{|A \cap B|}{|A \cup B|}$) as a good measure of similarity. Let us use $1 - J$ as a distance.

Does $1 - J$ obey triangular inequality for this set? [1]

(Hint Triangular inequality: $d(x, y) \leq d(x, z) + d(z, y)$)

- (a) YES
- (b) NO
- (c) Triangular inequality is not applicable for this problem.
- (d) Can not be computed.

[1] Advanced (optional): How do we show this for any general three sets?

Review Question - IV (one, none or more correct)

Consider the problem of feature transformation as

$$\mathbf{x}'_i = \mathbf{W}\mathbf{x}_i$$

If $\mathbf{x}_i \in R^2$ and $\mathbf{W}$ is 2×2 matrix with rank as 1, then

the new points $\mathbf{x}'_i$

- lie on a line in 2D
- are also R^2
- undefined
- One coordinate (dimension) of all the $\mathbf{x}'_i$ will be always the same
- all points in 2D will collapse into a single point. (i.e., $\mathbf{x}'_i = \mathbf{x}'_j$ for all i, j)
- none of the above.

Review Question - V (one, none or more correct)

In a TV Game show, a contestant selects one of three doors; behind one of the doors there is a prize,and behind the other two there are no prizes. After the contestant selects a door, the game-show host opens one of the remaining doors, and reveals that there is no prize behind it. The host then asks the contestant whether the you want to SWITCH their choice to the other unopened door, or STICK to their original choice.
[2]

What should we advise? What is the prob. of win if the candidate switch:

- (a) $\frac{1}{3}$ since all the doors are equally likely. Don't switch
- (b) $\frac{1}{2}$ since there are only two left, both are equally likely, no advantage in switching.
- (c) $\frac{2}{3}$. Prefer switching. Bayes says so.
- (d) $\frac{1}{3}$. Don't switching. Bayes says so.
- (e) None of the above.

---

[2] A very popular problem on internet from khan academy to mit lecture notes!. Appreciate the role of evidence, specially if the answer is not intuitive.