

Q8. Arrange in correct order

1. Preparing training data
2. Teach classifier
3. Output guesses on new data
4. Input unlabeled new data into model
5. Set aside hold-out set
6. Save model parameters
7. Verification with testing data

Correct answer: 1,5,2,6,7,4,3

Q12. Terminating conditions in neural network

1. Changes in weights are lesser than the threshold (I had written weights instead on changes in weights)
2. The percentage of tuples misclassified in the previous epoch is below some threshold
3. A prespecified number of epochs have expired.

Q15. Mapreduce**Working code written in mrjob for yelp business dataset:**

```
from mrjob.job import MRJob
from mrjob.protocol import JSONValueProtocol

class AvgStars(MRJob):
    INPUT_PROTOCOL = JSONValueProtocol
    def get_categories(self, _, record):
        if record['type'] == 'business':
            categories = record['categories']
            stars = record['stars']
            review_count = record['review_count']
            total_stars = stars * review_count
            for category in categories:
                yield [category, [total_stars, review_count]]

    def category_stars(self, category, tstars_reviews):
        tot_stars = 0
        tot_reviews = 0
        for sts, revs in tstars_reviews:
            tot_stars += sts
            tot_reviews += revs
        avg_stars = tot_stars / tot_reviews
        yield [category, avg_stars]

    def steps(self):
```

```

return [self.mr(self.get_categories, self.category_stars)]

if __name__ == '__main__':
    AvgStars.run()

```

Output:

```

"Accessories" 3.980769230769231
"Accountants" 4.678571428571429
"Active Life" 3.9925385508207594
"Acupuncture" 4.671052631578948
"Adult Education" 3.5
"Adult Entertainment" 2.914285714285714
"Adult" 3.880952380952381
"Advertising" 5.0
"Afghan" 4.005050505050505

```

Q7.3 Slice & Dice

Slice: Select a particular value in a dimension.

Eg. Revenue made by advertisements in one quarter, say Q1. The dimension here is time.

Dice: Subset of values of a dimension

Eg. Revenue made in Q1 and Q2 for advertisement types banner and text in pages home and search

Q7.4 Pivot

Rotate the data axes to provide alternate data presentation

Eg. In the slice example we had pages (home, search etc) in horizontal axis and revenue in vertical axis.

Pivot will swap these axes.

Q6 Star Schema

