Joanna Hsu

Midterm corrections

**#7 Data cubes**

7.3 Slice and dice

Slice: total views on all pages of banner advertisements on week 1

Dice: Total views of banner advertisements on home pages only on week 1

7.4 Pivot

-an axis change showing the same data.

-For example: initial query is for types of ads that can be shown on different types of pages. Then we pivot to show types of pages with different types of ads on.

**#8. Classifier correct order**

1. Prepare training data. Clean the data, set aside some to train the classifier.
2. Set aside hold-out set. Must save some to test with later!
3. Teach classifier. Various algorithms and methods possible.
4. Save model parameters.
5. Verification with testing data. How good is our model / classification rules?
6. Output guesses on new data If the accuracy is good enough, let’s try on new data.
7. Input unlabeled new data into model. This is the actual classification of unlabeled stuff!

**#11. Fixing gradient descent**

Our step size parameter is too large. We missed the minimum, and our errors are increasing instead of decreasing. We should decrease the step size.

**#13 Square or triangle**

where k=1. Triangle. The closest neighbor is a triangle.

Where k=5. Square. Of the 5 closest neighbors, 3 are squares (only 2 are triangles).

**#15. MapReduce**

MAPPER1 get\_categories

Input: line of data file

Output: key, value pairs: (category, [num\_stars, num\_reviews])

Example: restaurant, [5 stars, 6 reviews]

restaurant, [4 stars, 10 reviews]

Ice cream [5 stars, 6 reviews]

REDUCER1 aggregate\_by\_category

-input and output are both : (category, [num\_stars, num\_reviews])

Example: restaurant, [[5, 6],[4, 10]]

MAPPER do\_math

Input= output from reducer (category, [num\_stars, num\_reviews], [num\_stars, num\_reviews])

Example: restaurant, [[5, 6,],[4, 10]]

For each key, values pair, calculate the total number of stars and total number of reviews. Divide the total number of stars by total number of reviews.

Example output: restaurant, 4.375

REDUCER2 do\_nothing

Outputs the same input.