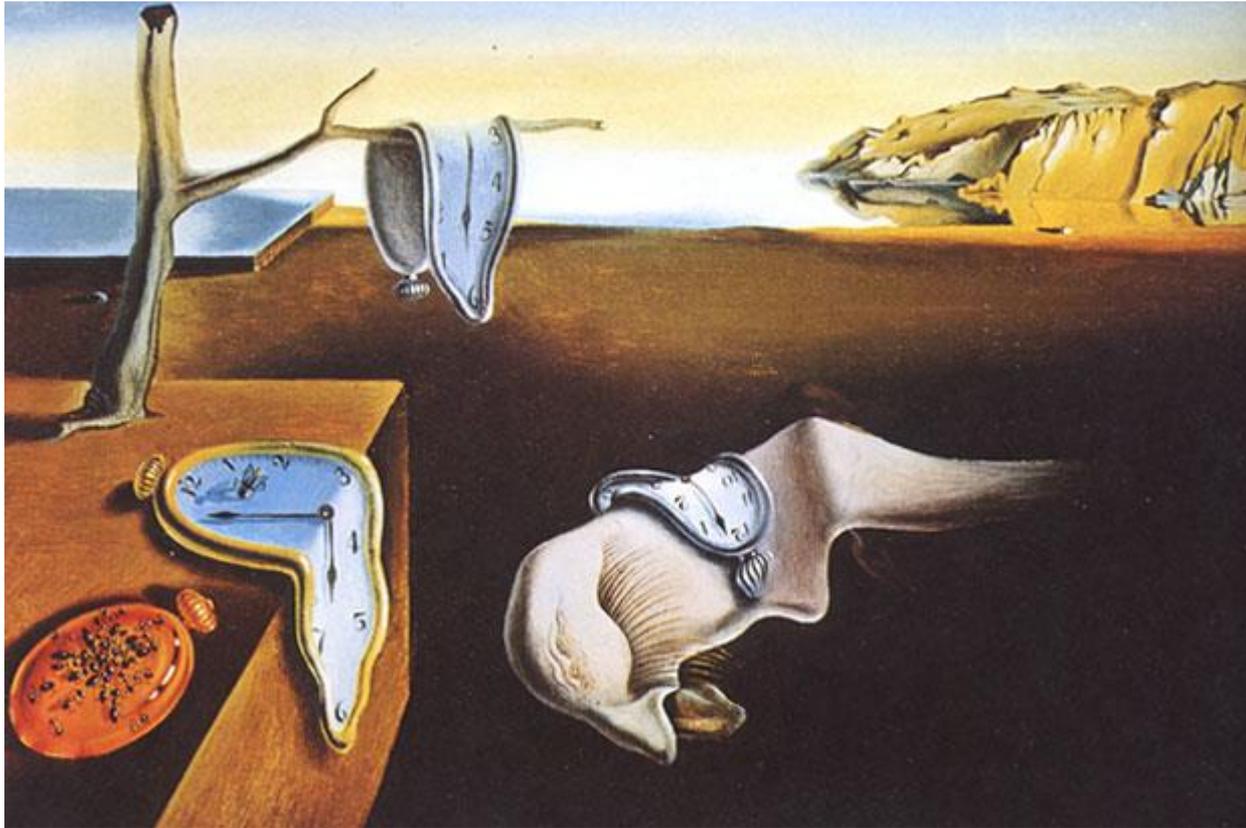# Midterm Post-mortem

Salvador Dali, The Persistence of Memory

CSC411/2515: Machine Learning and Data Mining, Winter 2018

Michael Guerzhoy and Lisa Zhang

Results

- Raw average: $\sim 50\%$
- Adjustment: $T \to 1.07T + 10$
  - T is the percentage grade (i.e., $raw/.90$)
  - Factor of 1.07 to make part of Q6 a bonus (so midterm is marked out of fewer marks
  - A bump by 10 points to account for the fact that there were too few "straightforward" questions that everyone would have gotten
  - Adjusted average: ~ 63.5
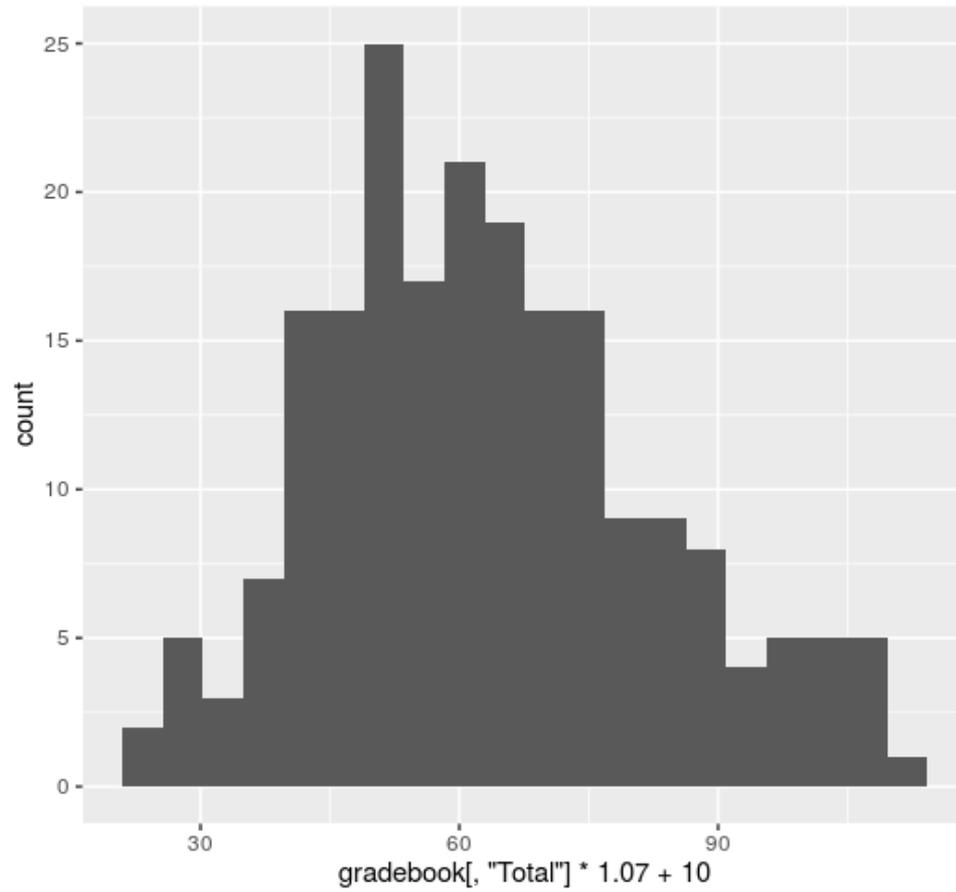
New marking scheme

- The larger of

40% Projects	40% Projects
10% Midterm	30% Midterm
50% Exam	30% Exam

- There is no target final average, so this can only help
- Needs to be approved by both the day and the night section
- Should allow people to recover from a poor showing on the midterm

Midterm results analysis

```
     MTQ1              MTQ2               MTQ3               MTQ4               MTQ5               MTQ6
Min.   : 3.00    Min.   : 0.000    Min.   : 0.000    Min.   : 0.000    Min.   : 0.000    Min.   : 0.000
1st Qu.:10.00    1st Qu.: 5.000    1st Qu.: 2.000    1st Qu.: 3.000    1st Qu.: 6.000    1st Qu.: 1.000
Median :12.00    Median : 7.000    Median : 4.000    Median : 6.000    Median : 8.000    Median : 3.500
Mean   :12.11    Mean   : 9.048    Mean   : 4.581    Mean   : 6.385    Mean   : 7.349    Mean   : 5.289
3rd Qu.:15.00    3rd Qu.:14.000    3rd Qu.: 7.000    3rd Qu.: 9.000    3rd Qu.:10.000    3rd Qu.: 7.500
Max.   :15.00    Max.   :20.000    Max.   :10.000    Max.   :15.000    Max.   :10.000    Max.   :20.000
```

- Q1 (multiple choice) and Q5 (capacity control) were pretty well done
- Q6 (EM algorithm) was the most challenging question

Doing better going forward

- The course is about 50/50 in terms of implementation and understanding the theory
  - Get an interview because you worked on cool ML projects, ace the interview because you show you understand ML, do great work because you can apply the theory in novel situations
- The study guide questions are basically homework
  - But most of them just ask you to make sure you understand what's going on in lecture
- There will be at least one project-related question on the exam

Doing better going forward

- The Study Guide is a fairly comprehensive list of questions that could be asked on the exam
- Most questions on the midterm were similar to study guide questions
  - Not even by design – it's just that the study guide is comprehensive
- **The study guide has "exercise"-type questions!**
  - For example, Q135 is basically the same as Q6 on the midterm, for a different model
  - There aren't a lot of "exercise"-type questions (maybe 25 up to now), so it's possible to do them all
- Work on improving the Google Docs solutions
  - This probably helps you more than it helps anyone else

Doing better going forward

- If a study guide question seems difficult, you discovered an area of the course that you don't completely understand

- Knowing the term "Guided Backprop" isn't super helpful – you need to understand the details of the algorithm

- It's easy to assume you understand a large formula on the slides because it e.g. just applies Bayes' rule. Don't assume that!
  - Make the formula "your own" by breaking it up into parts

Doing better going forward

- Understand the fundamentals
  - Probability: Likelihood, Maximum Likelihood, Maximum A-Posteriori, law of total probability, conditional expectation, generating data from models
  - Neural Networks
    - Multivariate chain rule
    - Most people seem to be doing well
  - Basis and dimension (for PCA)
- Understand the algorithms and methods
  - The big ones so far: Backprop, EM, learning decision trees, k-NN. Several minor algorithms like Guided Backprop

Doing better going forward

- Be clear on the terms used in the course
  - E.g., Likelihood
    - $P(data|\theta)$
  - E.g., weights (the strengths of the connections between neurons) vs. activations (what the neurons compute)
- Review math pre-requisites as needed
  - Understanding the derivative of the ReLU is with Q2 on the midterm

Historical CSC411 averages

- Course averages vary
  - Winter 2017 *final course average*: 77.8%, median: 80%.
- Need to be able to do things like Q4 (the Naïve Bayes question – straight from the slides and study guide) and Q3 (methods of visualizing a neuron in the 4-th layer) to do well

Doing better going forward

- Send me an email

Aside: PCA on grades

```
> prcomp(gradebook[c("MTQ1", "MTQ2", "MTQ3", "MTQ4", "MTQ5", "MTQ6")])
Standard deviations:
[1] 7.459249 3.885586 3.406712 2.833913 2.587123 2.250147

Rotation:
           PC1          PC2          PC3          PC4          PC5          PC6
MTQ1 0.1247829  0.005725397 -0.32276342 -0.37182455  0.804717381 -0.307240384
MTQ2 0.6362954 -0.757524578  0.08421956  0.11889929 -0.001822139  0.007169318
MTQ3 0.2511412  0.075372365 -0.35525089 -0.64985793 -0.269655701  0.556789345
MTQ4 0.3403633  0.306159606 -0.65965071  0.59212026 -0.021117596  0.065020712
MTQ5 0.1330223  0.039957924 -0.21265913 -0.26838915 -0.523367975 -0.767813612
MTQ6 0.6188242  0.570186285  0.53119176 -0.05152654  0.073160299 -0.042096768
```

- Reconstruction
  - $Q = \mu + \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$
- The best way to represent a student's grade without any parameters:

$$Q \approx \begin{pmatrix} 12.1 \\ 9.05 \\ 4.58 \\ 6.39 \\ 7.35 \\ 5.29 \end{pmatrix}$$

Aside: PCA on grades

```
> prcomp(gradebook[c("MTQ1", "MTQ2", "MTQ3", "MTQ4", "MTQ5", "MTQ6")])
Standard deviations:
[1] 7.459249 3.885586 3.406712 2.833913 2.587123 2.250147

Rotation:
            PC1          PC2          PC3          PC4          PC5          PC6
MTQ1 0.1247829  0.005725397 -0.32276342 -0.37182455  0.804717381 -0.307240384
MTQ2 0.6362954 -0.757524578  0.08421956  0.11889929 -0.001822139  0.007169318
MTQ3 0.2511412  0.075372365 -0.35525089 -0.64985793 -0.269655701  0.556789345
MTQ4 0.3403633  0.306159606 -0.65965071  0.59212026 -0.021117596  0.065020712
MTQ5 0.1330223  0.039957924 -0.21265913 -0.26838915 -0.523367975 -0.767813612
MTQ6 0.6188242  0.570186285  0.53119176 -0.05152654  0.073160299 -0.042096768
```

- Reconstruction
  - $Q = \mu + \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$
- The best way to represent a student's grade with one parameter:

$$Q = \begin{pmatrix} 12.1 \\ 9.05 \\ 4.58 \\ 6.39 \\ 7.35 \\ 5.29 \end{pmatrix} + \lambda \begin{pmatrix} 0.12 \\ 0.63 \\ 0.25 \\ 0.34 \\ 0.12 \\ 0.61 \end{pmatrix}$$

All grades increase together