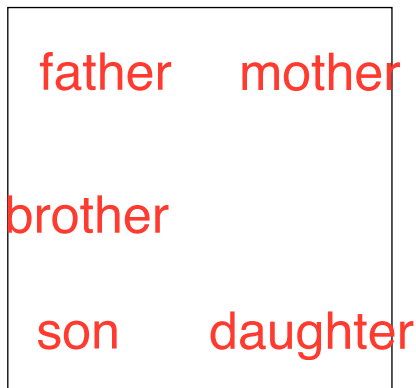1. Can we use accuracy as loss in SGD? Why or why not?

2. Why is "surprise" a good intuition for log-loss (categorical cross-entropy) but not MSE? What are the units of each?
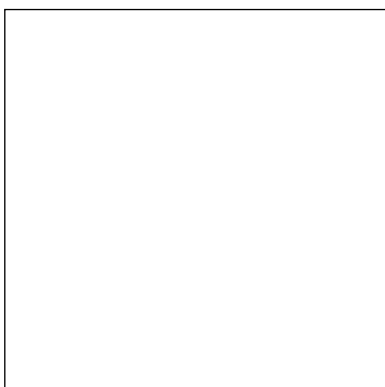
accuracy is not differentiable

"surprise" - of the correct class
MSE units: squared target units
log-loss: units are bits / nats

3. *embeddings warm-up*: (a) Place the following words in the space below so that similar words are closer together: *father, mother, son, daughter*. (b) Once you've done that, now try to add *brother*. Explain why the flat paper makes this hard; what would you do if you had another dimension to work with?

father   mother

brother

son   daughter

Before you leave, pick a couple of these questions to react to:

1. What was the most important concept from today for you?
2. What was the muddiest concept today?
3. How does what we did today connect with what you've learned before or with the objectives for this course?
4. What would you like to review or clarify (from today or otherwise) next time we met?
5. What are you curious, hopeful, or excited about?