# CS 541-B Artificial Intelligence: Final Exam

Instructor: Jie Shen

12/14/2022, 18:30 – 21:00 EST

**Instructions:**

- Closed book exam;
- Always give your answer and explain it;
- 20 points per problem, totally 110 points (20 * 5 + 10).

**0.** Your name. (10 pts)

**1.** Talk about your understanding of a problem that has been discussed in the class.

**2.** Random projection and principal component analysis (PCA) are two prominent dimension reduction approaches. Discuss when one performs better than the other.

**3.** Give an example where active learning leads to exponential saving of labeling cost; give another example that it does not.

**4.** Suppose we have the following data from $n$ patients:

| | Age | Weight | Height | Gender | Blood Pressure | $\cdots$ | Sharp Pain |
|---|---|---|---|---|---|---|---|
| Patient 1 | $z_{11}$ | $z_{12}$ | $z_{13}$ | $z_{14}$ | **?** | $\cdots$ | $z_{1m}$ |
| Patient 2 | $z_{21}$ | $z_{22}$ | $z_{23}$ | $z_{24}$ | $z_{25}$ | $\cdots$ | $z_{2m}$ |
| Patient 3 | $z_{31}$ | $z_{32}$ | $z_{33}$ | $z_{34}$ | **?** | $\cdots$ | $z_{3m}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| Patient $n$ | $z_{n1}$ | $z_{n2}$ | $z_{n3}$ | $z_{n4}$ | $z_{n5}$ | $\cdots$ | $z_{nm}$ |

where some entries in the column Blood Pressure are missing (represented by the symbol "?"), and other columns are fully observed.

- Give an algorithm to estimate these missing values based on the current data matrix.
- How can we estimate the missing values if some of the observed entries are corrupted?

**5.** Suppose we have the following data from $n$ patients:

| | Age | Weight | Height | Gender | Blood Pressure | $\cdots$ | Sharp Pain |
|---|---|---|---|---|---|---|---|
| Patient 1 | **?** | $z_{12}$ | $z_{13}$ | $z_{14}$ | **?** | $\cdots$ | $z_{1m}$ |
| Patient 2 | **?** | $z_{22}$ | **?** | $z_{24}$ | $z_{25}$ | $\cdots$ | **?** |
| Patient 3 | $z_{31}$ | **?** | $z_{33}$ | **?** | **?** | $\cdots$ | $z_{3m}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| Patient $n$ | **?** | $z_{n2}$ | $z_{n3}$ | **?** | $z_{n5}$ | $\cdots$ | **?** |

where for each column and each row there are some missing entries (represented by the symbol "?"). Give an algorithm to estimate all the missing values.