

# Machine problem 1

- Applying KNN to MNIST dataset
- Using three test protocols to evaluate the performance
  - Training/test split: 50000 training examples + 10000 test examples
  - Training/validation/test split: 40000 training examples + 10000 validation examples + 10000 test examples
  - 5-fold cross-validation and 10-fold cross validation (average and standard deviation)
- Due in two weeks (Sep 2nd before class)

# What shall be submitted?

- Your implementation
  - What language you choose to implement the algorithm?
  - How do you search for the K nearest neighbors in the feature vector space? Exhaustive search? Or smarter one (hint: k-d tree)?
  - Your algorithm's complexity to predict on one test example?
- Your results
  - Test errors under different test protocols
  - How performance changes when K varies? Why?
- Submitting your source code
- Electric report in PDF or Word formats
- Where: cap5610ucf@gmail.com.