Introduction to Machine Learning

Final Project

- Requirements
- Timeline
- Test and Report

- Groups of 1-3 students
  - Report should list the contribution of each student explicitly
  - Each student should record his/her contribution

Application

- Image data

Minimum requirements

- 2 options for data normalization
- PCA with 2 options for number of components (dimensions)
- Data mapping using clustering (2 options)
- Feature Selection (2 options)
- 4 different types of classifiers
- Ensemble methods: *Bagging* and *AdaBoost*
- Use k-fold cross validation with k=4 for all validations (nested 4-fold if needed)
- Use *Pipelines* and *GridSearch*
- Base your algorithm/parameter selection on accuracy, AUC of ROC, and F1-measure

Minimum requirements

- DO NOT
  - Combine all options and parameters in one giant GridSearch!
  - Use any algorithm/technique that was not covered in class
- DO
  - Consider few options at a time
  - Analyze the results
  - Justify your next set of options

Report 1

- Due Nov 25 (there will be penalty for late submission: **5 points off per day**)
- Worth **30 points**, graded based on
  - Experiments, results, and analysis
  - Discussion/justification of remaining experiments
  - This report should include about 50% of all your experiments

Final report

- Due Dec 4 (no late submission)
  - *NO REPORTS WILL BE ACCEPTED OR GRADED AFTER 04/24*
- Worth 40 points, graded based on
  - Experiments, results, considered options, etc.
  - Discussion of the most important parameters that affect the results
  - Performance of the classifier (based on cross-validation of the provided data)
  - Analysis of the results
    - Visualization of correct samples, confused samples, etc.
    - Possible justification for misclassified samples

Test (of new images)

- By Dec 2nd, you will be provided with the features of 30 test images (no labels and no images).
  - You need to test these images with your best model and submit a csv file that has the labels of the 20 images.
  - We will score these test images
- Worth max of 30 points.
  - Depending on all available results (including ours), we will grade your accuracy as
    - **Excellent** (30 pts)
    - **Good** (20 pts)
    - **Average** (10 pts)
    - **Don't make sense** (e.g. all assigned to the same class) or code doesn't run (0 pts)
- Bonus points
  - Best: + 15points Second: +10points Third: + 5points