Some information about the midterm

Guangliang Chen

October 15, 2018

Outline

- Machine learning terminology
- Dimensionality reduction techniques
  - PCA (unsupervised)
  - LDA (supervised)
- Classifiers
  - Instance-based: kNN, NLC
  - Bayes classifiers: (kNN), LDA/QDA, Naive Bayes

Midterm

- Theoretical tools:
  - Linear algebra (matrix multiplication, Rayleigh quotient, SVD),
  - Probability and statistics (multivariate normal and Bayes rule), and
  - Multivariable calculus (method of Lagrange multipliers)
- Programming (homework assignments)

Midterm Exam

Date: Oct. 17, Wed, in class

- Covers all material up to Bayes classifier (inclusively)
- No books or notes will be allowed.
- 5 Questions (concept, computing by hand, and proof)

Midterm

Conceptual questions:

- Orthogonal / symmetric / positive (semi)definite matrices, Rayleigh quotient
- Training/test data, classifier, cross validation
- Training/validation/test error, confusion matrix, decision boundary
- kNN classification

Midterm

Hand calculations:

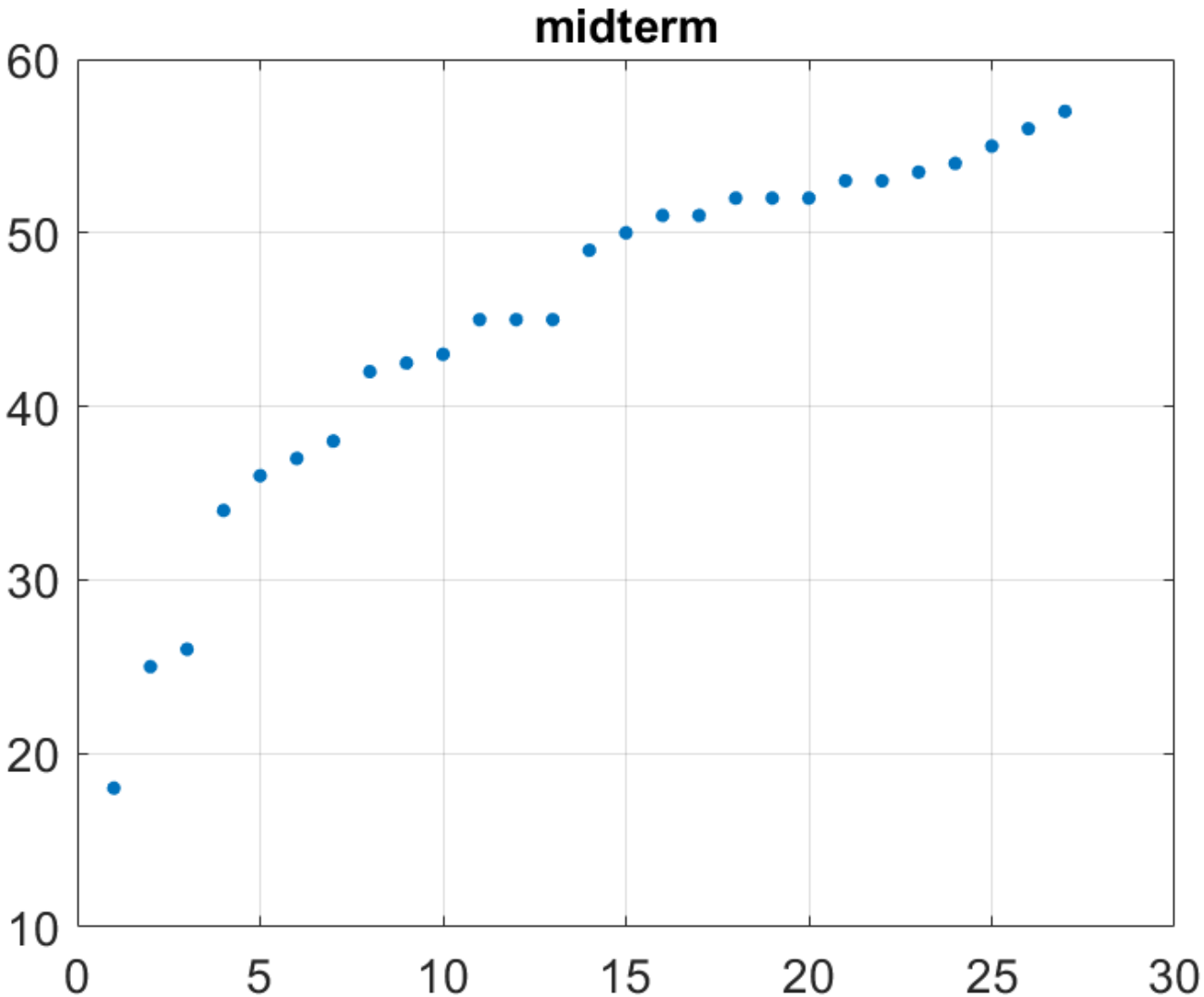
- Eigenvalue decomposition of a symmetric matrix
- SVD (+terminology)
- PCA (+interpretation)
- Two-class LDA

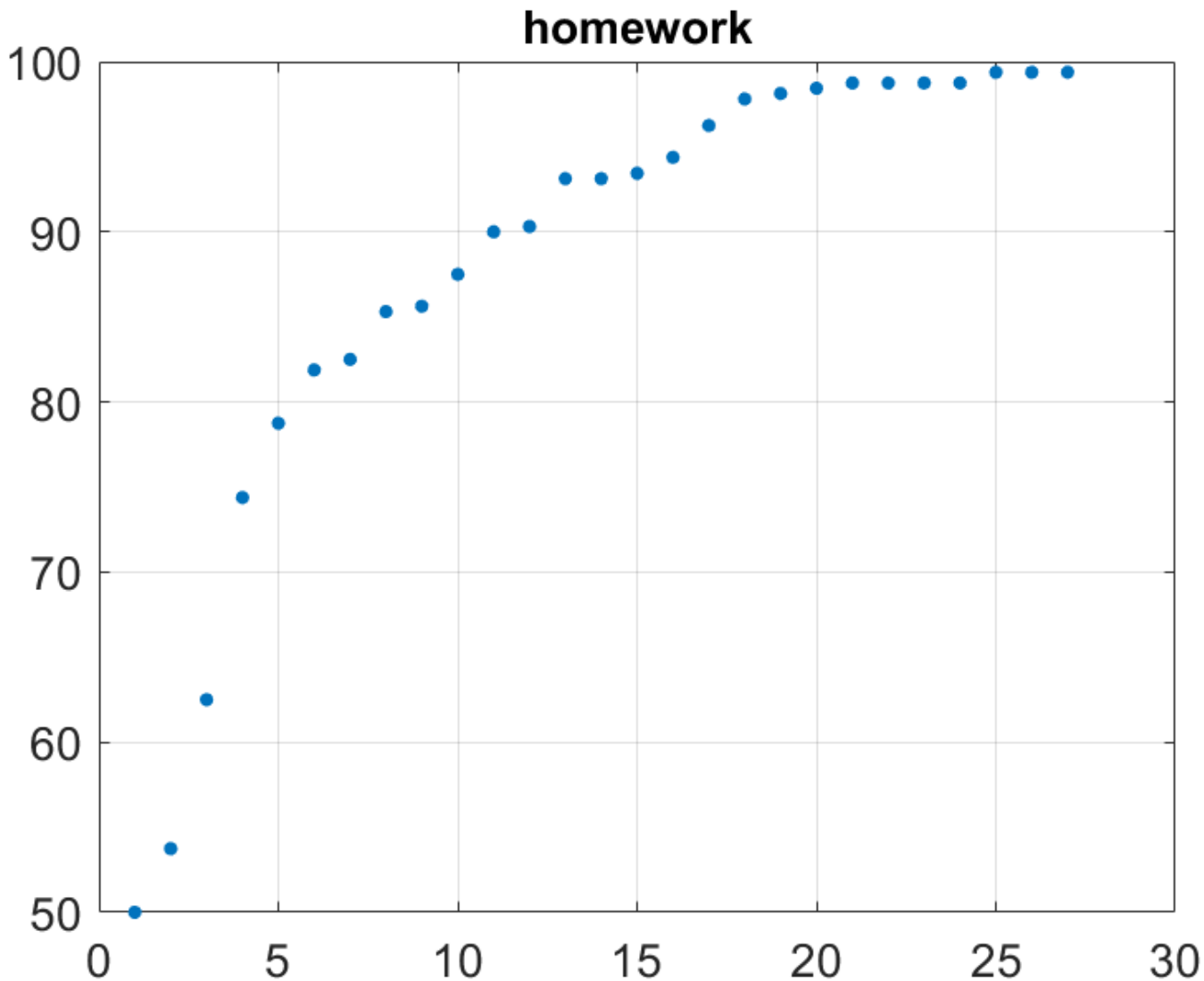
Midterm

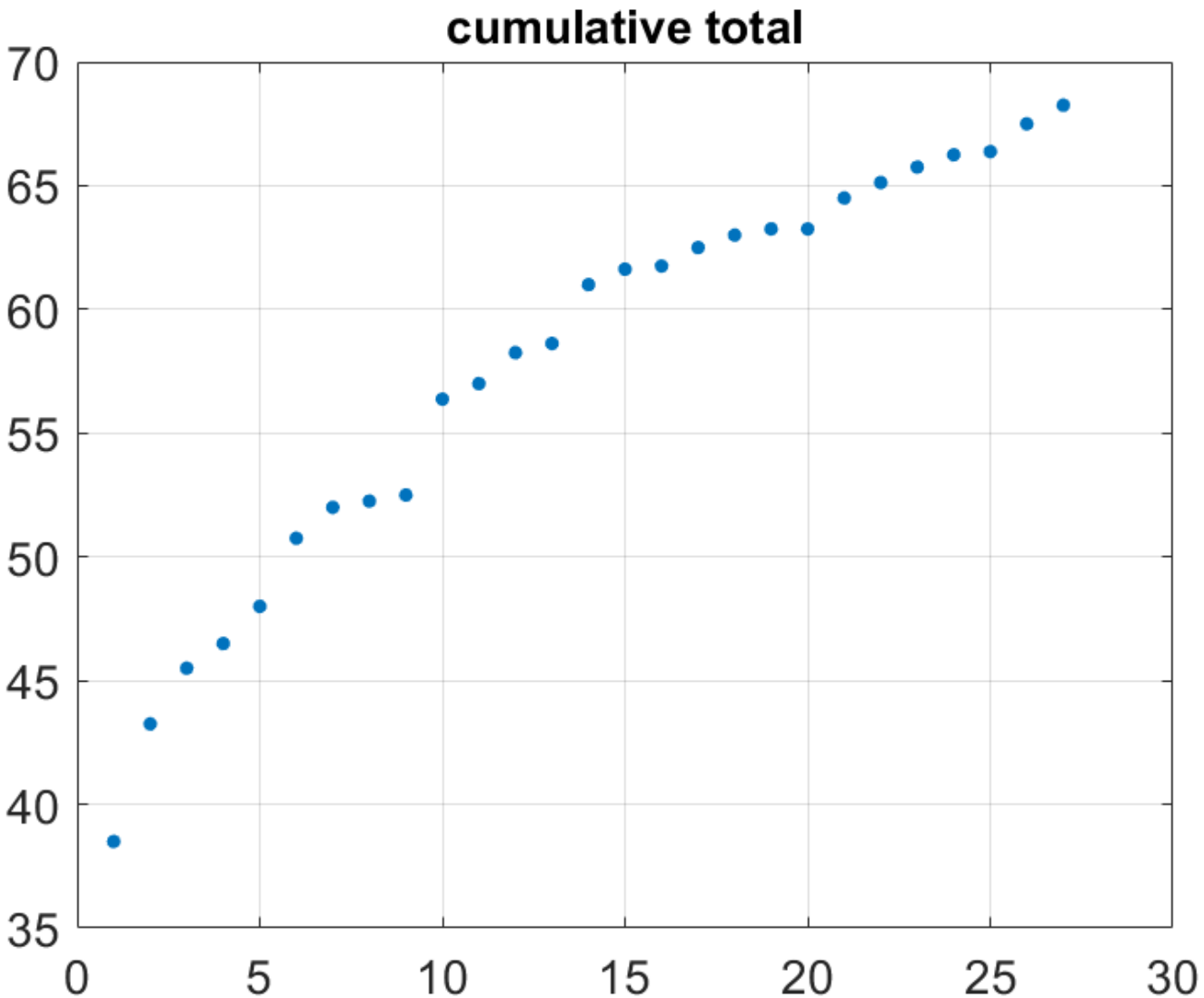
Derivation/proofs

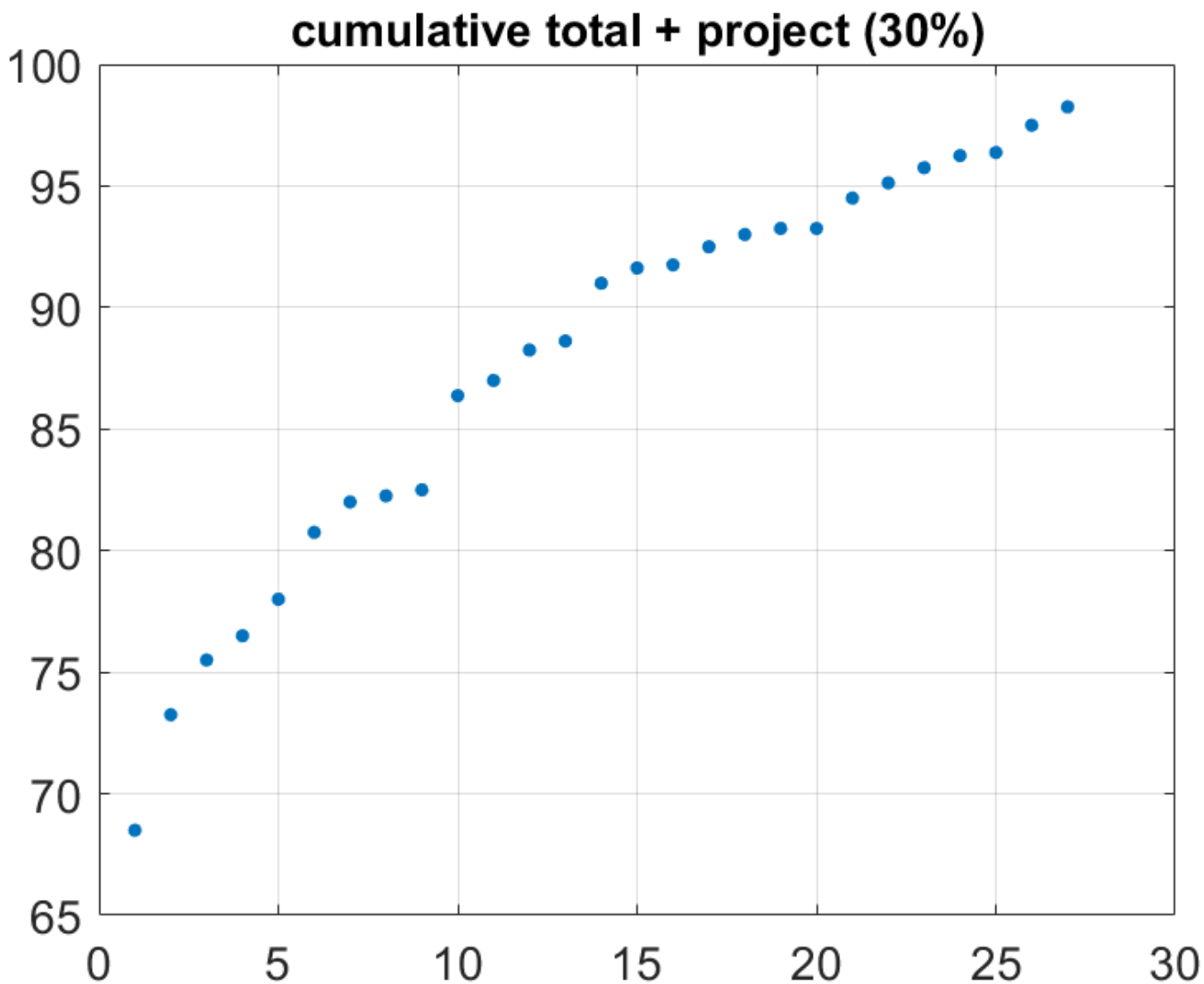
- PCA result
- Two-class LDA result
- Bayes classifier (with 1D class distributions)

Midterm




homework
100
90
80
70
60
50
0
5
10
15
20
25
30


cumulative total
70
65
60
55
50
45
40
35
0
5
10
15
20
25
30


cumulative total + project (30%)
100
95
90
85
80
75
70
65
0
5
10
15
20
25
30

List of possible final projects

- "Interesting" data used in HW1-Q5
- kNN variants (probabilistic kNN, etc.)
- Singularity of LDA (PCA, regularization, pseudoinverse)
- Kernel kNN / PCA / LDA (first transform nonlinear data to another space)
- 2D PCA / LDA (handle images as matrices directly, no vectorization)
- Other dimensionality reduction approaches (such as MDS)
- Categorical data reduction and/or classification
- Text data reduction and/or classification