Math 251: Statistical & Machine Learning Classification

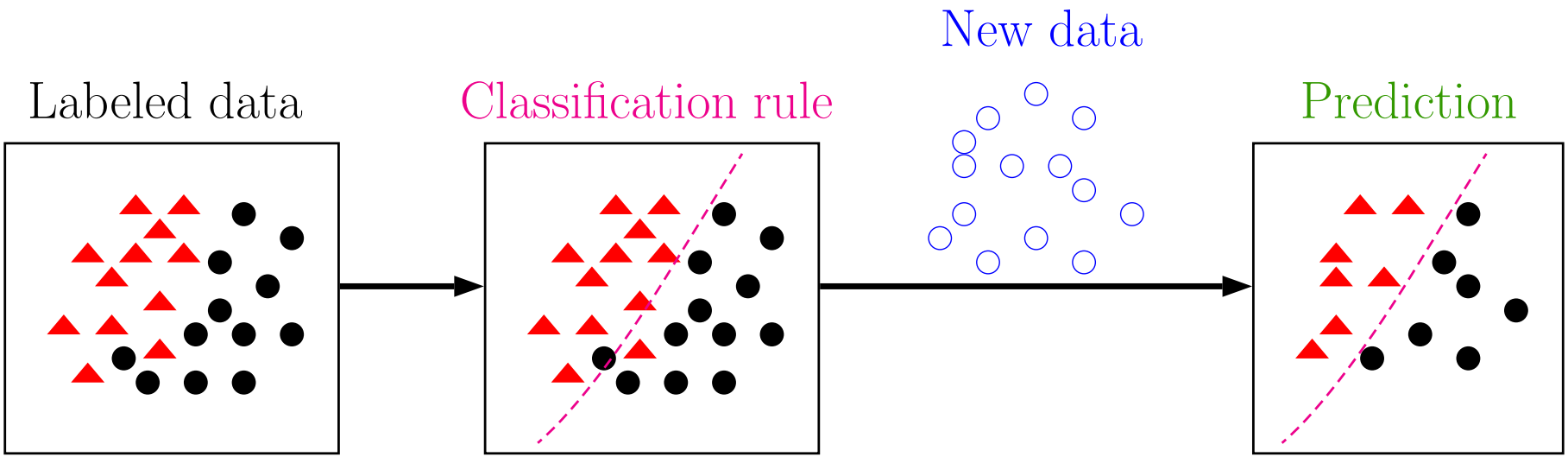
Course summary and project information

Dr. Guangliang Chen

Department of Math & Statistics
San José State University

Content and level of this course

An introduction to the machine learning field of classification, the task of assigning labels to new data based on a given set of labeled data.



Methods covered in this course

- Dimensionality reduction: PCA, LDA
- Instanced-based classifiers: kNN and variants
- Bayes classifiers: LDA/QDA, Naive Bayes
- Logistic regression
- Support vector machine (soft, hard, kernel, multiclass extension)
- Ensemble methods: trees, bagging, random forest, and boosting
- Neural networks

Balancing theory and programming

- Theory tools: Linear algebra, multivariable calculus, probability and statistics, optimization
- Programming tools: R, Matlab, Python

Further learning opportunities

- CMPE 258 Deep Learning
- MATH 280 Graduate Individual Studies
- MATH 298 Special Study

Teaching the course to you is learning to me

- Fall 2015: Math 203 CAMCOS ⟵ training
- Spring 2016: Math 285 Classification with Handwritten Digits ⟵ validation
- Fall 2018: Math 251 Statistical & Machine Learning Classification ⟵ testing?

Stochastic gradient ascent?

- Fall 2015: Math 203 CAMCOS ⟵ initialization
- Spring 2016: Math 285 Classification with Handwritten Digits ⟵ epoch 1
- Fall 2018: Math 251 Statistical & Machine Learning Classification ⟵ epoch 2

Important reminder:
Please complete the course evaluation by next Monday (if you haven't). Your helpful feedback is greatly appreciated and will be used to improve the course next time.

Project information

- Project presentations:
  - Monday (Dec 10), 9-10:15am ⟵ Submit your slides to Canvas by 5pm, Sunday (Dec 9) to secure a spot on a first-come, first-serve basis.
  - Wednesday (Dec 12), 7:15-9:30am ⟵ Submit your slides to Canvas by 5pm, Tuesday (Dec 11)
- Project report: due Wed (Dec 12)

Presentations must be ...

- 7 minutes long each (no exception) $\longleftarrow$ practice, practice, practice
- at proper level (you should only assume the background of this course; anything beyond must be clearly explained)
- complete (title, introduction, problem definition, method, experimental setting, results, discussions, conclusions, and acknowledgments/references)
- clear (use sufficiently large font size in table/plot, add title/label/legend to your plots, put talking points on slides, provide enough detail)
- high-level (focus on idea, leave detailed derivation/math to report)

Project reports must be ...

- at least 5 pages long $\longleftarrow$ no upper bound on length
- complete (title, introduction, problem definition, method, experimental setting, results, discussions, conclusions, and acknowledgments/references)
- clear (generally free of English mistakes, provide all the necessary details including derivation/math)
- typed in LaTex (preferred) or Word!

Questions?