UCLA Samueli Computer Science

CS M146 Discussion: Week 10

PCA, HMM, Final Review

Junheng Hao

Friday, 03/12/2021

UCLA
Engineer Change.

Roadmap

- Announcement
- PCA
- HMM
- Q & A

UCLA
Engineer Change.

Announcement

- **There is no quiz in Week 10.**
- **Problem set 4** released on CCLE, submission on Gradescope.
  - Please assign pages of your submission with corresponding problem set outline items on GradeScope.
  - You need to submit code and the results required by the problem set
  - Due on **today 11:59pm PST, Mar 12 (Friday)**
- **Final Exam: March 15 (Next Monday)**
  - “Quiz-like” exam, submission through GradeScope

Late Submission of PS and final exam will NOT be accepted!

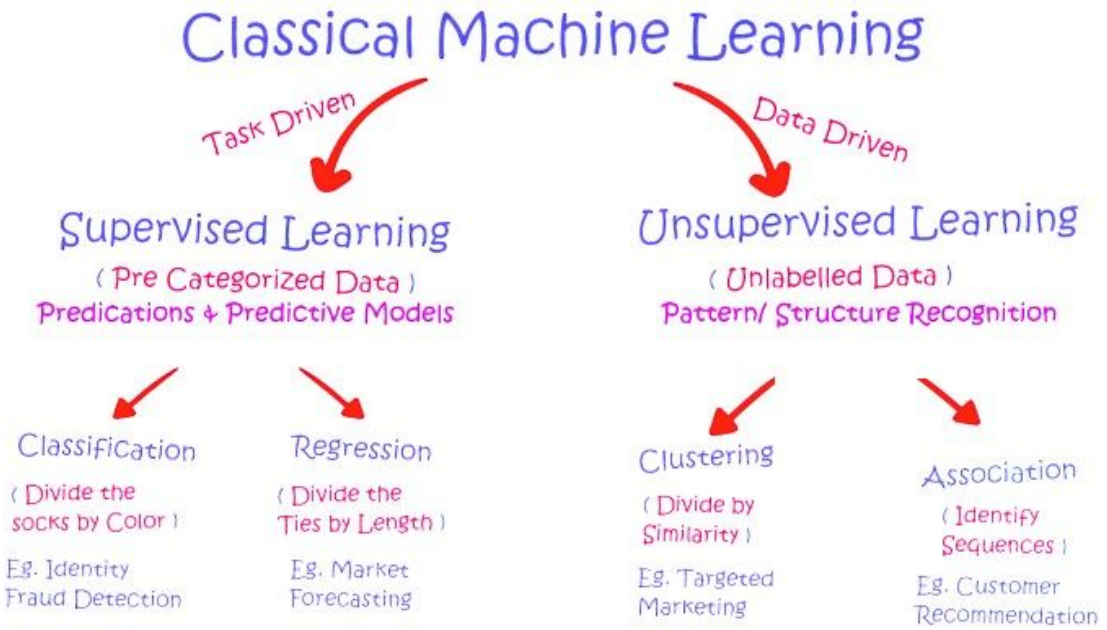
Updates: Final Exam

- Open book and open notes, on GradeScope: “quiz”-like exam
- Start attempting the exam from **8:00 am PST on March 15**; Submit your exam before **8:00am PST March 16** (No extensions). → 24h time window
- **You must start before 5:00am PST March 16 to use the full 3 hours. No late submission time.**
- Exam duration: **3 hours** (time limit after start the exam)
- Type: True/false and multiple choice questions (free text boxes are given for justification)
- The instructors will be available to provide clarifications on CampusWire (visible for everyone) from 8:00am-11:00am on March 15. Later questions on Campuswire may not be answered.
- Some calculations are expected.

MUST READ: Official post about final exam on Campuswire:
https://campuswire.com/c/GB5E561C3/feed/437

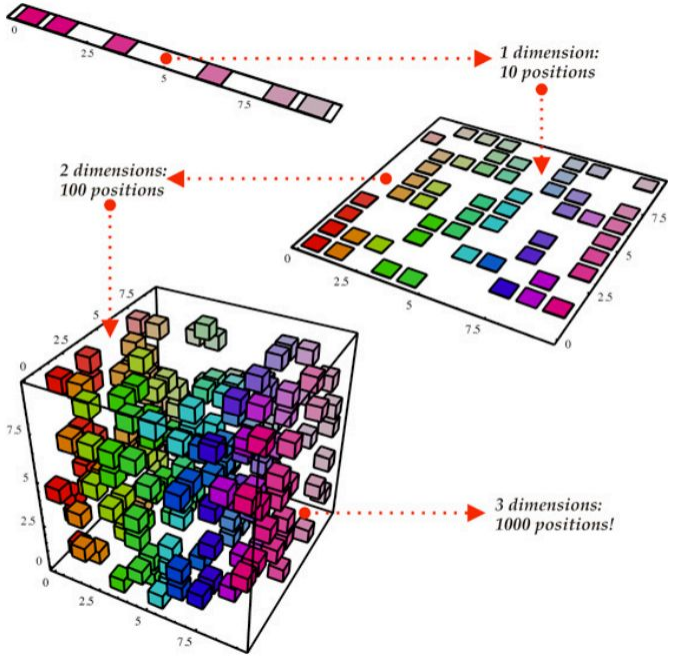
Supervised learning / Unsupervised learning

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

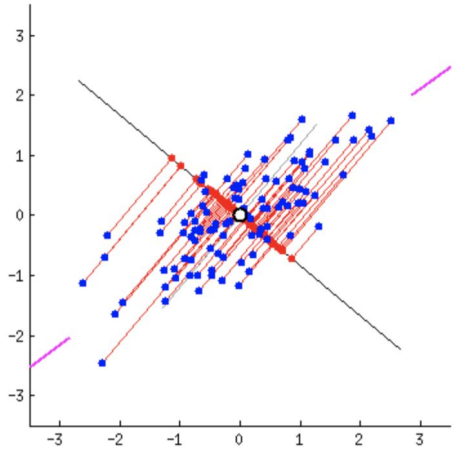


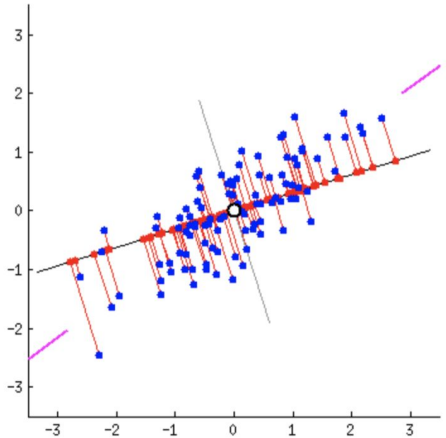
Large feature space and dimension reduction

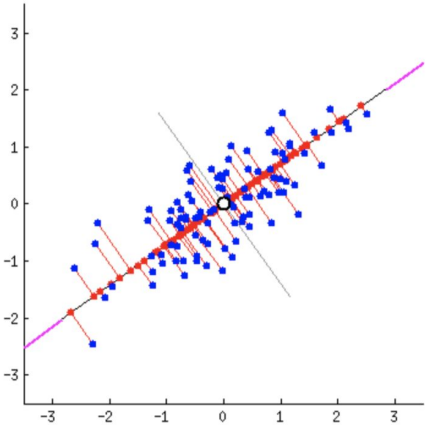
- Disadvantages of having a large feature space
  - More data is required
  - Redundant features and more noise → Model overfitting
  - Algorithm's simplicity and fewer assumptions [Occam's razor]
- Straightforward dimensionality reduction
  - Feature elimination
  - Feature extraction



PCA



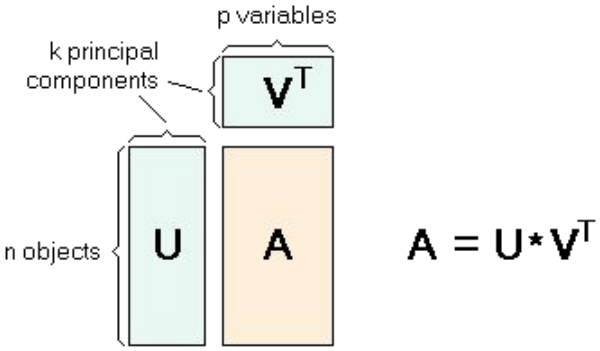


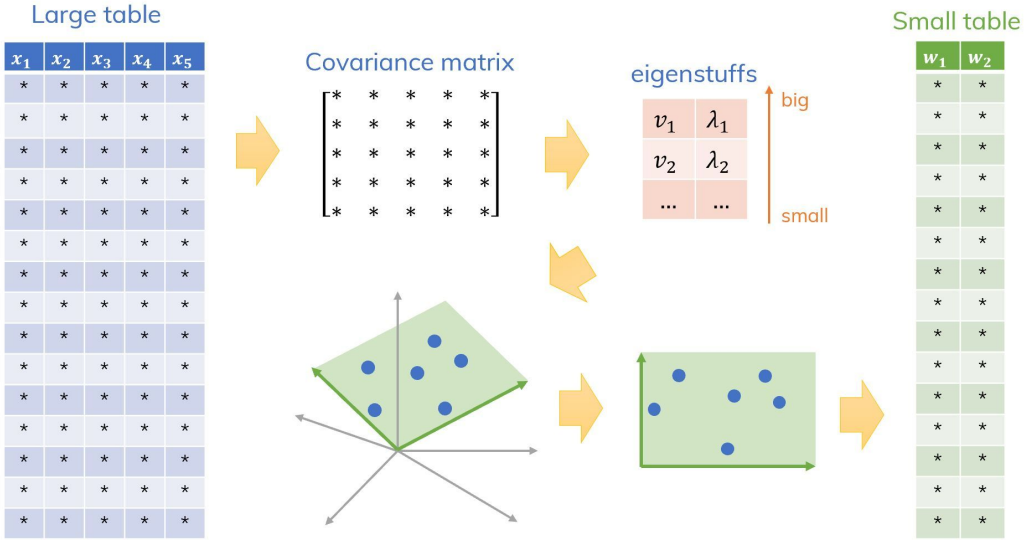


PCA: Formulation

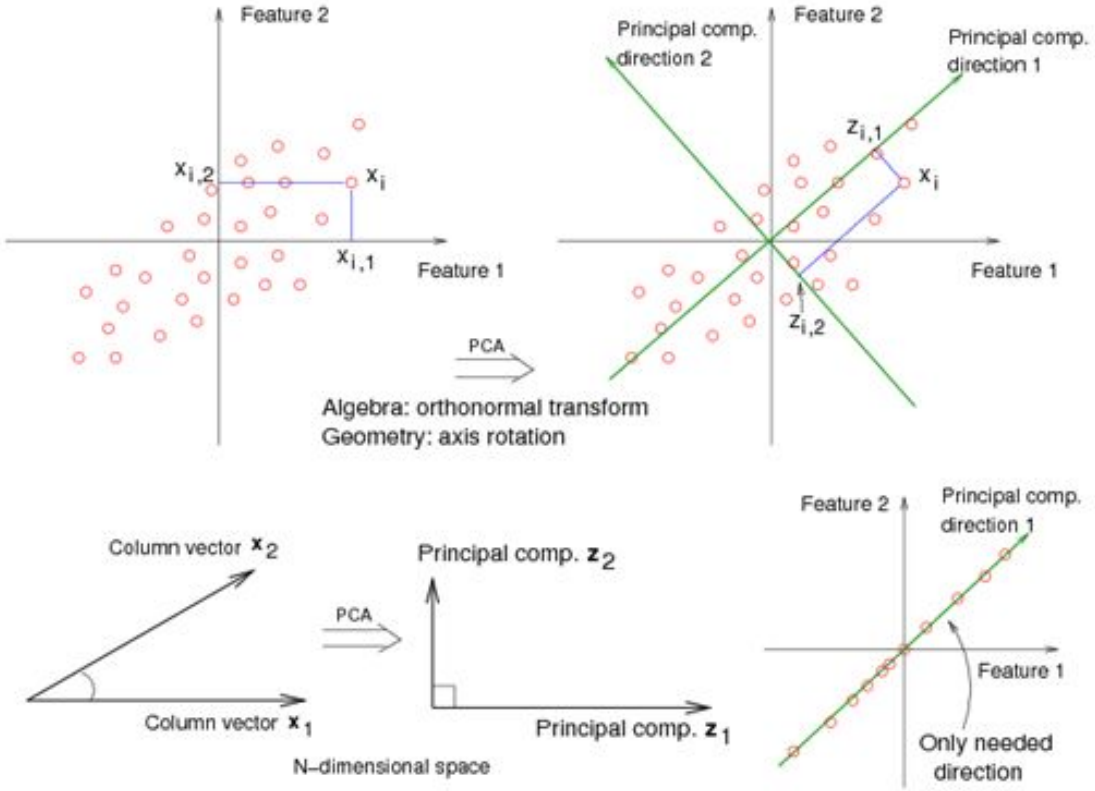
- Dimension reduction as matrix decomposition

$$\begin{bmatrix} \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{P} \end{bmatrix}$$





PCA: Geometric Interpretation



Credit: https://online.stat.psu.edu/stat508/lesson/6/6.3

PCA: Formulation in 1-dim (k=1)

- Demo shown in whiteboard

PCA Approach: Summarization

Steps:

- Take the whole dataset consisting of ***d***-dimensional samples
- Compute the d-dimensional mean vector (i.e., the means for every dimension of the whole dataset)
- Compute the covariance matrix of the whole data set
- Compute eigenvectors and corresponding eigenvalues
- Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a ***d×k*** dimensional matrix (where every column represents an eigenvector)
- Use this ***d×k*** eigenvector matrix to transform the samples onto the new subspace.

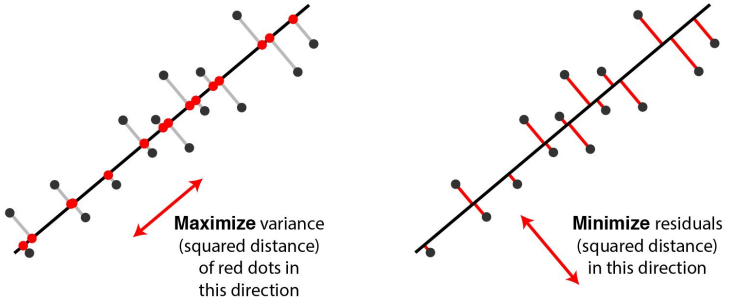
Credit: https://sebastianraschka.com/Articles/2014_pca_step_by_step.html

UCLA
Engineer Change.

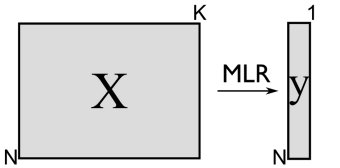
PCA: Example

- Demo shown at: https://sebastianraschka.com/Articles/2014_pca_step_by_step.html

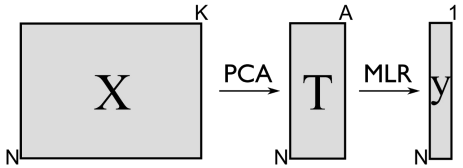
PCA vs Linear Regression



Multiple linear regression



Principal component regression



HMM: Concepts

- Markov Process: the next state depends on the current state

$$P(x_{t+1}|x_1, \ldots, x_t) = P(x_{t+1}|x_t)$$

$$P(x_1, \ldots, x_t) = P(x_1)P(x_2|x_1) \ldots P(x_t|x_{t-1})$$

- Initial probability

$$\pi_i = P(X_1 = i)$$

- Transition probability

$$q_{ij} = P(X_{t+1} = i|X_t = j)$$

- Emission symbols

$$e_i(b) = P(Y_t = b|X_t = i)$$

UCLA Engineer Change.

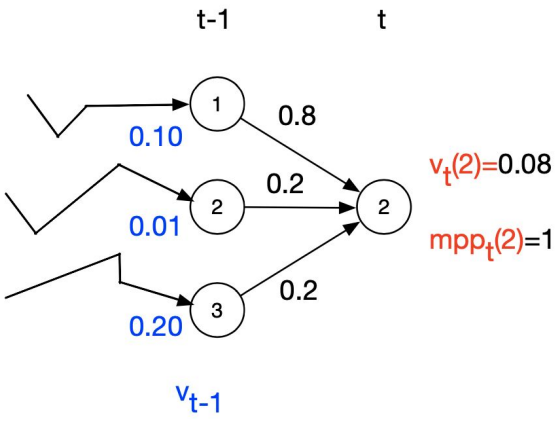
HMM: Computing probability of a sequence

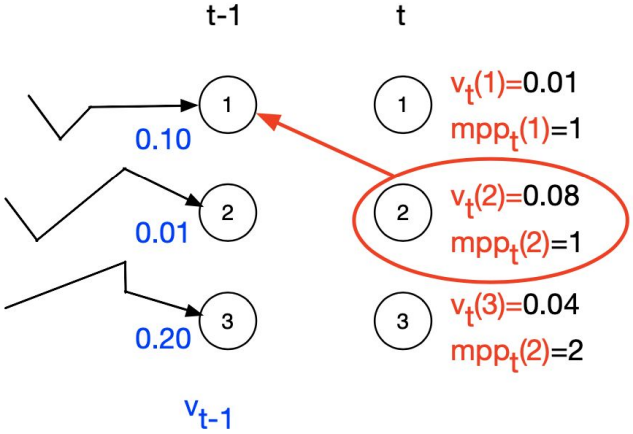
- Assume uniform probability of starting in each states and transition probability matrix

$$\boldsymbol{Q} = \begin{pmatrix} 0.5 & 0.1 & 0.0 \\ 0.3 & 0.0 & 0.4 \\ 0.2 & 0.9 & 0.6 \end{pmatrix}$$

HMM: Most probable path

- Problem: Given a sequence of observations, what is the most probable sequence of hidden states?
- Solution: Viterbi Algorithm





HMM: Viterbi/Dynamic Programming

- The most probable path with last two states (l, k) is the most probable path with state l at time (t–1) followed by a transition from state l to state k and emitting the observation at time t.

$$v_{t-1}(l)P(X_t = k|X_{t-1} = l)P(y_t|X_t = k)$$
$$= v_{t-1}(l)q_{lk}e_t(y_t)$$

- Maximization process

$$v_t(k) = \max_l v_{t-1}(l)q_{kl}e_t(y_t)$$

$$mpp_t(k) = l^*$$
$$l^* = \arg\max_l v_{t-1}(l)q_{kl}e_t(y_t)$$

HMM: Viterbi Algorithm Example [Link]

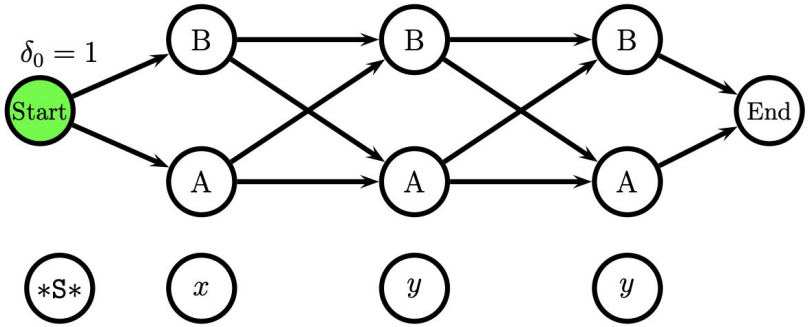
Transition probability distribution

Current	Next: A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

Emission probability distribution

State	Word: *S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

Given a sequence of < ***S***, ***x, y, y,*** ***E**>

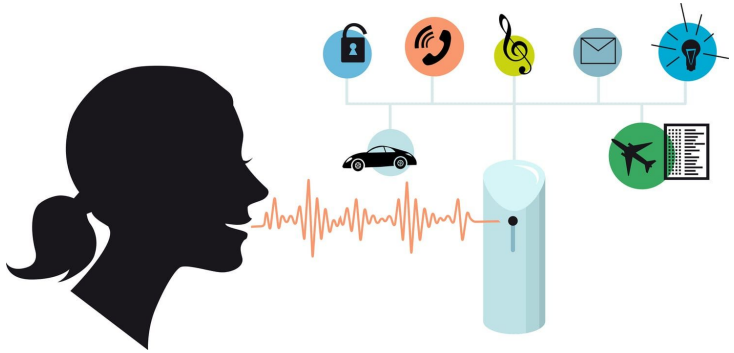


*HMM: Learning Parameters

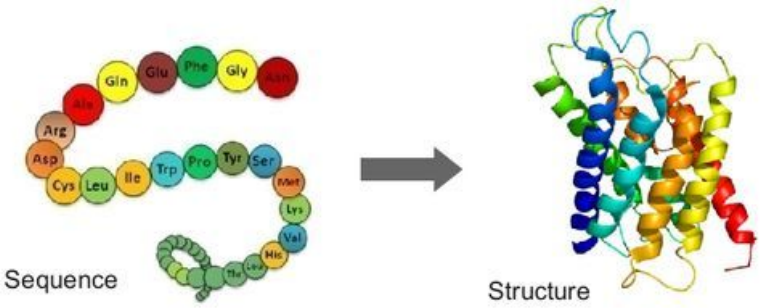
- Solution: Baum–Welch algorithm

HMM: Applications and restrictions

Speak Recognition



Protein Structure Prediction



UCLA
Engineer Change.

CSM 146: Summary

	Supervised Learning	**Unsupervised Learning**
Model	Decision tree, kNN Neural nets,	K-means, GMM PCA, HMM
Loss Function	0/1, square, hinge, exponential, cross entropy (log)	
Optimization	MLP, MAE, SVM (dual problem, constrained) Gradient descent (batch / stochastic) EM algorithms	
Theory	PAC learning, VC-dimension	
Others	Convexity/concavity, hyperparameters, overfitting and underfitting, inductive biases, regularizations	

UCLA
Engineer Change.

Advanced Classes in AI/ML+

1. CS174A: Introduction to Computer Graphics (Prof. Asish Law, etc)
2. CS247: Advanced data mining (Prof. Yizhou Sun)
3. CS240: Big data seminar / Graph neural network (Prof. Yizhou Sun / Wei Wang)
4. CS260: Machine learning algorithms (Prof. Quanquan Gu/ Prof/ Cho-Jui Hsieh)
5. CS22X: Algorithms in Bioinformatics / Advanced Computational Genetics / Computational Methods in Medicine (Prof. Sankararaman / Prof. Eskin)
6. CS263: Natural language processing (Prof. Kai-Wei Chang / Nanyun Peng)
7. CS269: Seminars in deep learning foundations / natural language processing, etc. (Prof. Quanquan Gu / Kai-Wei Chang / Nanyun Peng)

Other courses are taught in EE, Stats departments:

1. ECE 236B/C: Convex Optimization (Prof. Vandenberghe)
2. ECE 239AS Reinforcement Learning Theory and Applications / Neural networks (Prof. Lin F. Yang / Jonathan Kao)

End-of-quarter Congratulations!

Professors: you made it to the end of the semester. congratulations

Students:

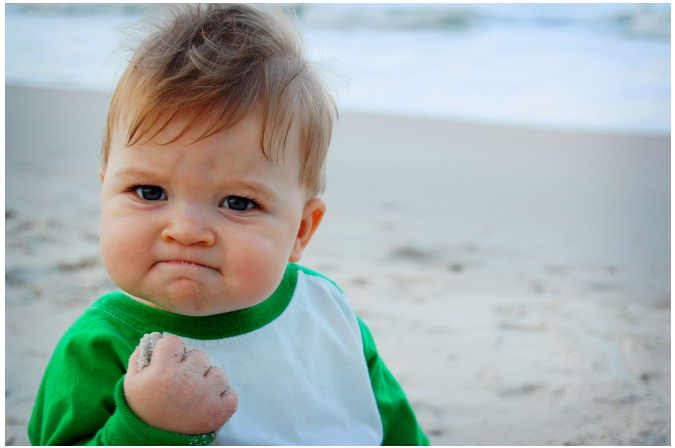
UCLA Samueli Computer Science

Thank you

for learning with us in winter 2021.

Good luck

as certified ML experts!



UCLA Samueli Computer Science

Evaluation of Instruction

Reminder: You have until **Saturday, March 13 8:00 AM PST** to complete confidential evaluations for CSM146 and Dis 1C (Junheng).

Math Backup: Eigendecomposition