

# CS145: INTRODUCTION TO DATA MINING

## Final Review

---

**Instructor: Yizhou Sun**

[yzsun@cs.ucla.edu](mailto:yzsun@cs.ucla.edu)

December 6, 2017

# Learnt Algorithms

	Vector Data	Set Data	Sequence Data	Text Data
Classification	Logistic Regression; Decision Tree; KNN; SVM; NN			Naïve Bayes for Text
Clustering	K-means; hierarchical clustering; DBSCAN; Mixture Models			PLSA
Prediction	Linear Regression GLM*			
Frequent Pattern Mining		Apriori; FP growth	GSP; PrefixSpan	
Similarity Search			DTW	

# Final Exam

---

- Time
  - 12/13, 11:30am-1:30pm
- Location
  - Royce Hall 362
- Policy
  - Closed book exam
  - You can take two “reference sheets” of A4 size, i.e., one in addition to the midterm “reference sheet”
  - You can bring a **simple** calculator

# Content to Cover

- All the content learn so far

- ~ 20% before midterm
- ~ 80% after midterm

	Vector Data	Set Data	Sequence Data	Text Data
Classification	Logistic Regression; Decision Tree; KNN; SVM; NN			Naïve Bayes for Text
Clustering	K-means; hierarchical clustering; DBSCAN; Mixture Models			PLSA
Prediction	Linear Regression GLM*			
Frequent Pattern Mining		Apriori; FP growth	GSP; PrefixSpan	
Similarity Search			DTW	

# Type of Questions

---

- Similar to Midterm
  - True or false
- Conceptual questions
- Computation questions