

BREAST CANCER DIAGNOSIS

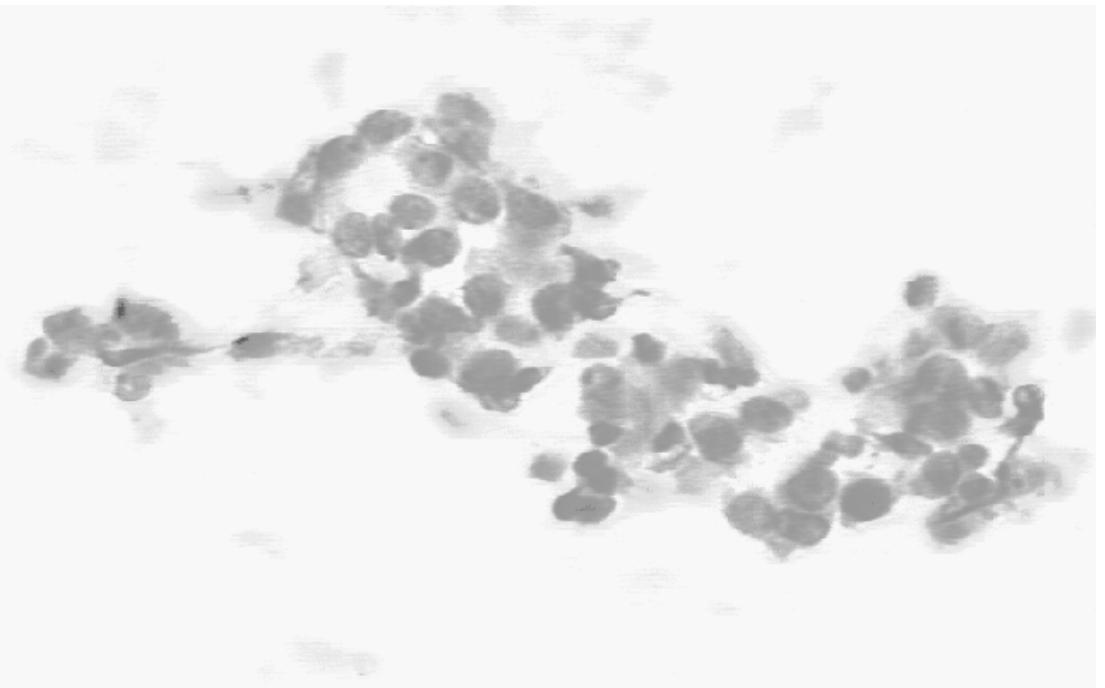
Group 6:

Feng Qingduo, Shen Zhentao, Sun Wenjia, Lin Jialiang



CATALOG

- Preface
- Dataset
- Task Goal
- Preliminary Plan
 - Stage 1
 - Stage 2



PREFACE

- COVID-19 (Corona Virus Disease 2019) has become a global pandemic.

- Total Confirmed: **937,091**
- Total Death: **47,231**

(up to 10:00AM, April 2)



- Breast Cancer is a serious issue
 - From WHO (World Health Organization)'s estimation :
 - Impact **2.1 million** women each year
 - About **627,000** women died from breast cancer in 2018



DATASET

- Wisconsin Diagnostic Breast Cancer

- From UC Irvine Machine Learning Repository ([UCI-MLR](#))
 - Can also be found in [Kaggle](#)

- 32 Attributes: ID, diagnosis, and 30 real-valued input attributes

- Diagnosis = {benign, malignant}
 - 10 features obtained from digitized image of FNA (fine needle aspirate) of a breast mass.



- 1)radius
- 2)texture
- 3)perimeter
- 4)area
- 5)smoothness
- 6)compactness
- 7)concavity
- 8)concave points
- 9)symmetry
- 10)fractal dimension



TASK GOAL

- Classification Task

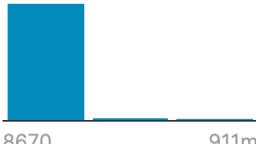
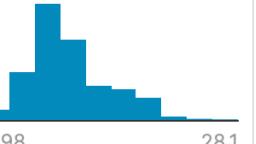
- determine whether a new record data is related with benign breast cancer or malignant breast cancer

- Supervised learning with labels



PLAN

- Stage 1
 - data exploration & data preprocessing
 - divide training-set & test test
 - feature selection
 - build classifier
 - analyze the result

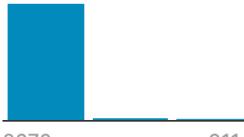
	id	diagnosis	radius_mean
	ID number	The diagnosis of breast tissues (M = malignant, B = benign)	mean of distances from center to points on the perimeter
		B 63% M 37%	 6.98 28.1
1	842302	M	17.99
2	842517	M	20.57
3	84300903	M	19.69
4	84348301	M	11.42



PLAN

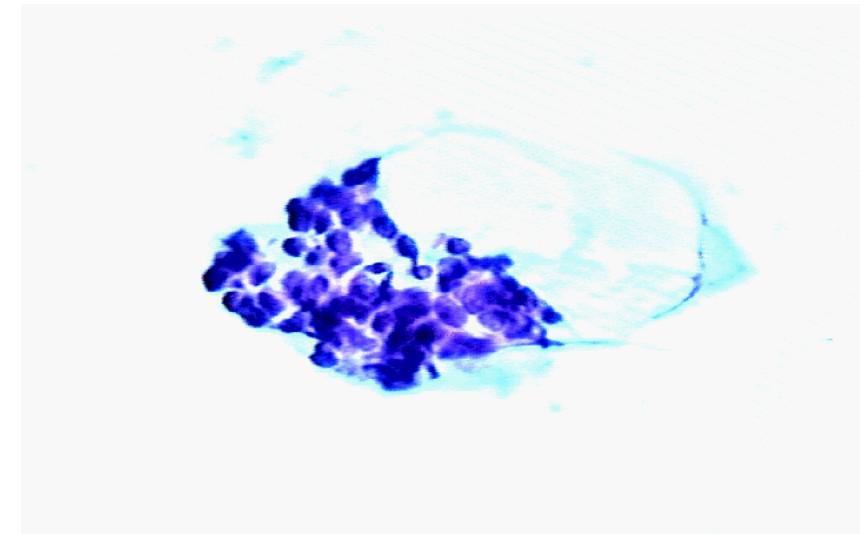
- Stage 1

- data exploration & data preprocessing
- divide training-set & test test
- feature selection
- build classifier
- analyze the result

	id	diagnosis	# radius_mean
	ID number	The diagnosis of breast tissues (M = malignant, B = benign)	mean of distances from center to points on the perimeter
	 8670	B 63%	 6.98
	911m	M 37%	28.1
1	842302	M	17.99
2	842517	M	20.57
3	84300903	M	19.69
4	84348301	M	11.42

- Stage 2

- extract features from original images
- image process to get 10 features in WDBC
- explore more features
- ...



REFERENCE

- COVID-19 data: JHU covid-19 data
(<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6/>)
- Breast Cancer data: WHO website (<https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>) and American Cancer Society (<https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html>)
- Wisconsin Diagnostic Breast Cancer dataset: (<https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>)
- Breast Cancer Image: (<ftp://cs.wisc.edu/math-prog/cpo-dataset/machine-learn/WDBC/>)



THANK YOU

Q&A

Group 6:

Feng Qingduo, Shen Zhentao, Sun Wenjia, Lin Jialiang

