

# DS-DC-13

# LECTURE NOTES

*Jim Simpson*  
*Data Scientist, Sotera*

LESSON 08

# LESSON 8

## LEARNING OBJECTIVES

- DEFINE CLASS LABELS & CLASSIFICATION
- BUILD K-NEAREST NEIGHBORS CLASSIFIER
- EVALUATE USING CLASSIFICATION METRICS
  - ACCURACY
  - MISCLASSIFICATION ERROR
  - CONFUSION MATRIX

# MACHINE LEARNING

↓  
Predict a  
continuous  $y$   
i.e.  $y$  is real-valued

Linear  
Regression

LESSON  
6

↓  
Predict a  
categorical  $y$   
i.e.  $y$  is a class label

k-Nearest  
Neighbors

Logistic  
Regression

Random  
Forest

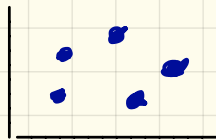
LESSON  
8

# CURSE OF DIMENSIONALITY

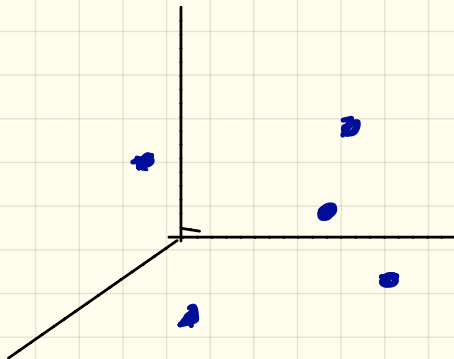
1D



2D



3D

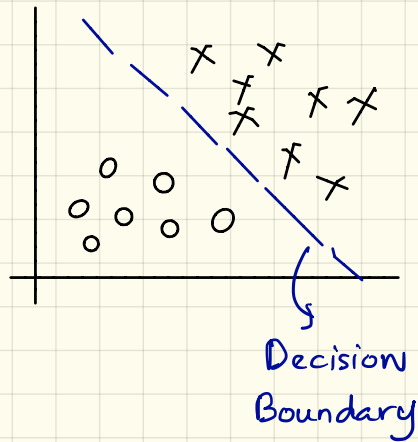


As  
dimensionality  
increases

Available  
data  
becomes  
sparse



# CONFUSION MATRIX



		PREDICTED CLASS		
		O	X	
ACTUAL CLASS	O	TP 91%	FN 9%	→ False Neg Type II error
	X	FP 8%	TN 92%	

↓  
False Positive  
Type I error