EE5263

Deep Learning

Spring, 2021

Lecture 2 - Introduction to Machine Learning

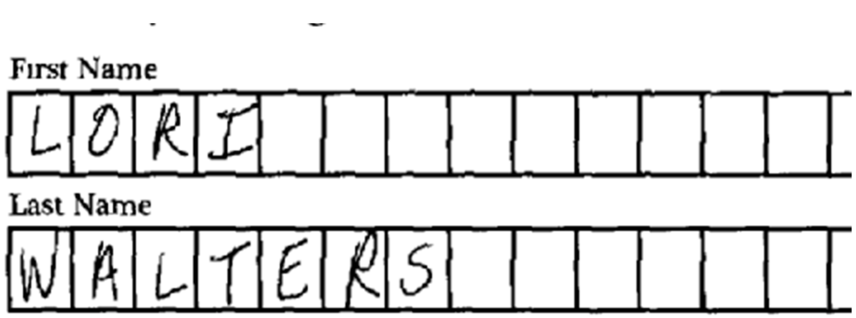
Instructor: **Yufei Huang**, PhD, *University of Texas at San Antonio*

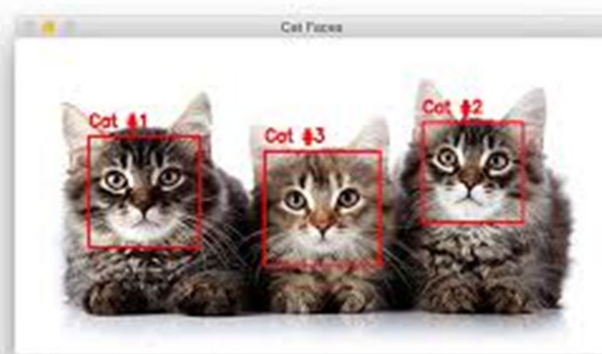
UTSA

Machine Learning

– Machine Learning is the ability to teach a computer without explicitly programming it

– Examples are used to train computers to perform tasks that would be difficult to program

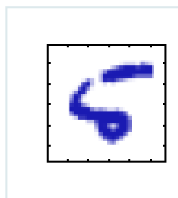




Applications of Machine Learning

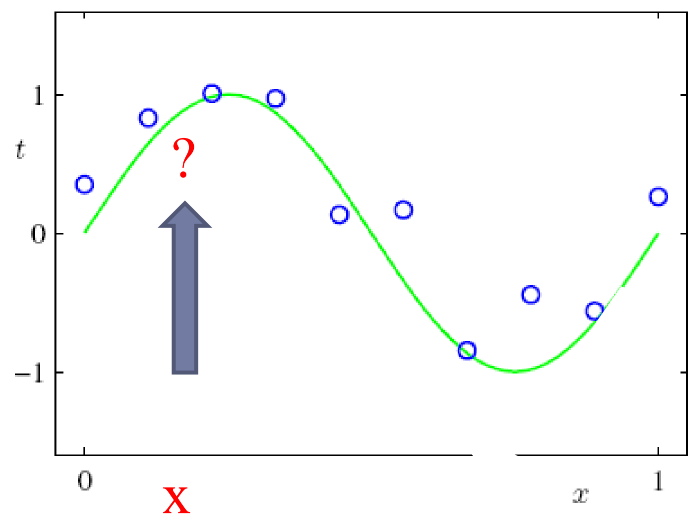
- Handwriting Recognition
  - convert written letters into digital letters
- Language Translation
  - translate spoken and or written languages (e.g. Google Translate)
- Speech Recognition
  - convert voice snippets to text (e.g. Siri, Cortana, and Alexa)
- Image Classification
  - label images with appropriate categories (e.g. Google Photos)
- Autonomous Driving
  - enable cars to drive

Hand-written digit recognition (classification)

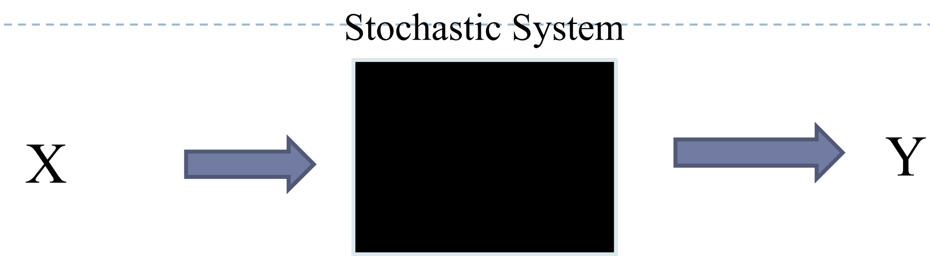


?

Curve fitting (regression)

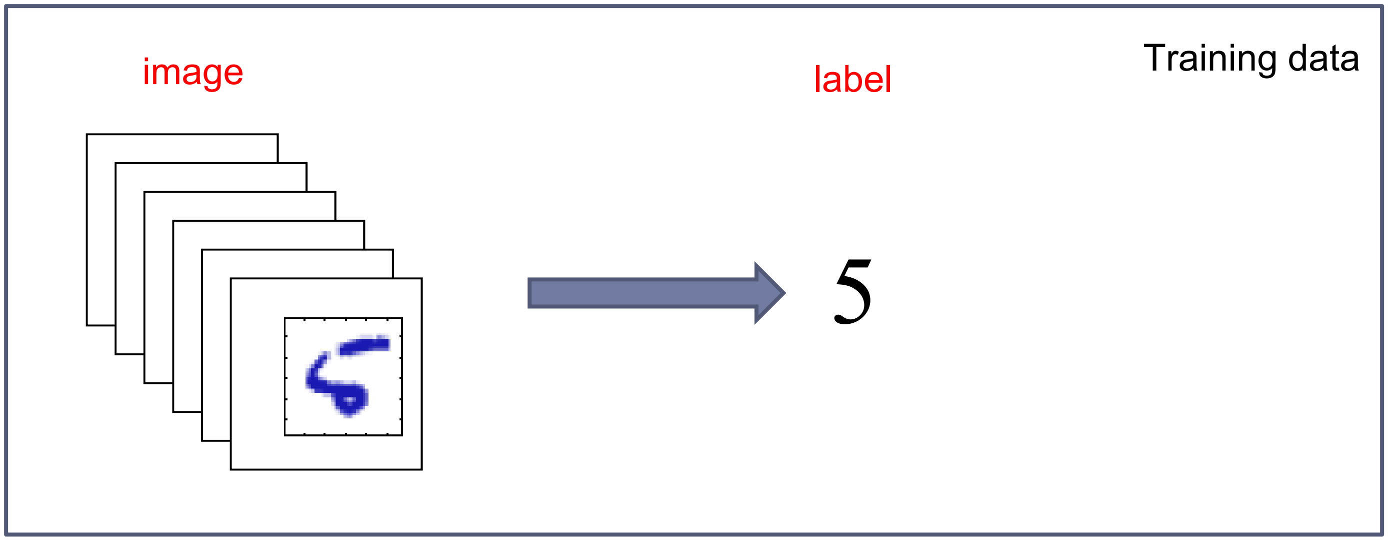


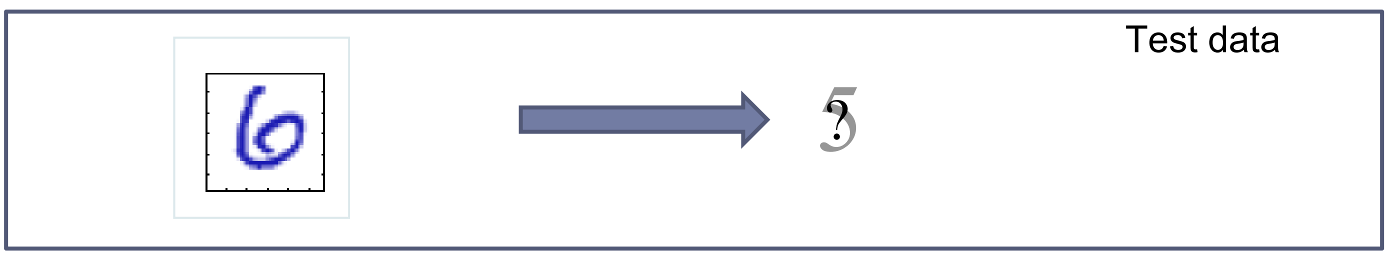
Types of Machine Learning



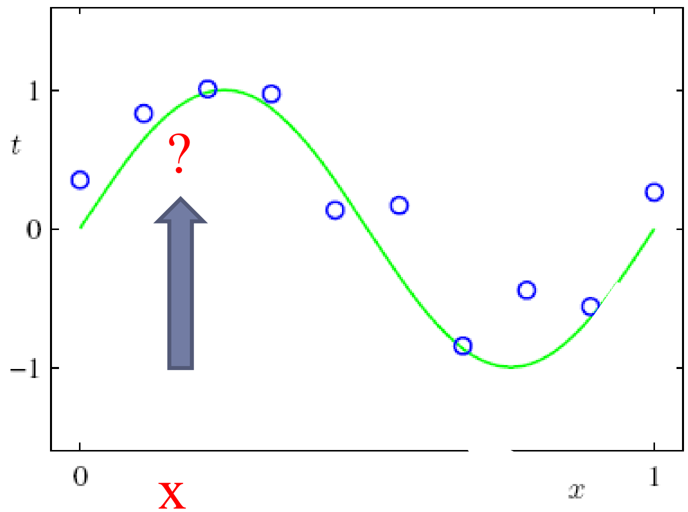
- Supervised learning
  - Given a set of input and output (label), or training data, the goal is to determine the label of a new input
  - Regression and Classification
- Unsupervised learning
  - Given a set of data, the goal is to determine the structure/categories in data
  - Clustering
- Reinforcement Learning
  - Training data is unlabeled
  - System receives feedback for its actions
  - Goal is to perform better actions

Classification

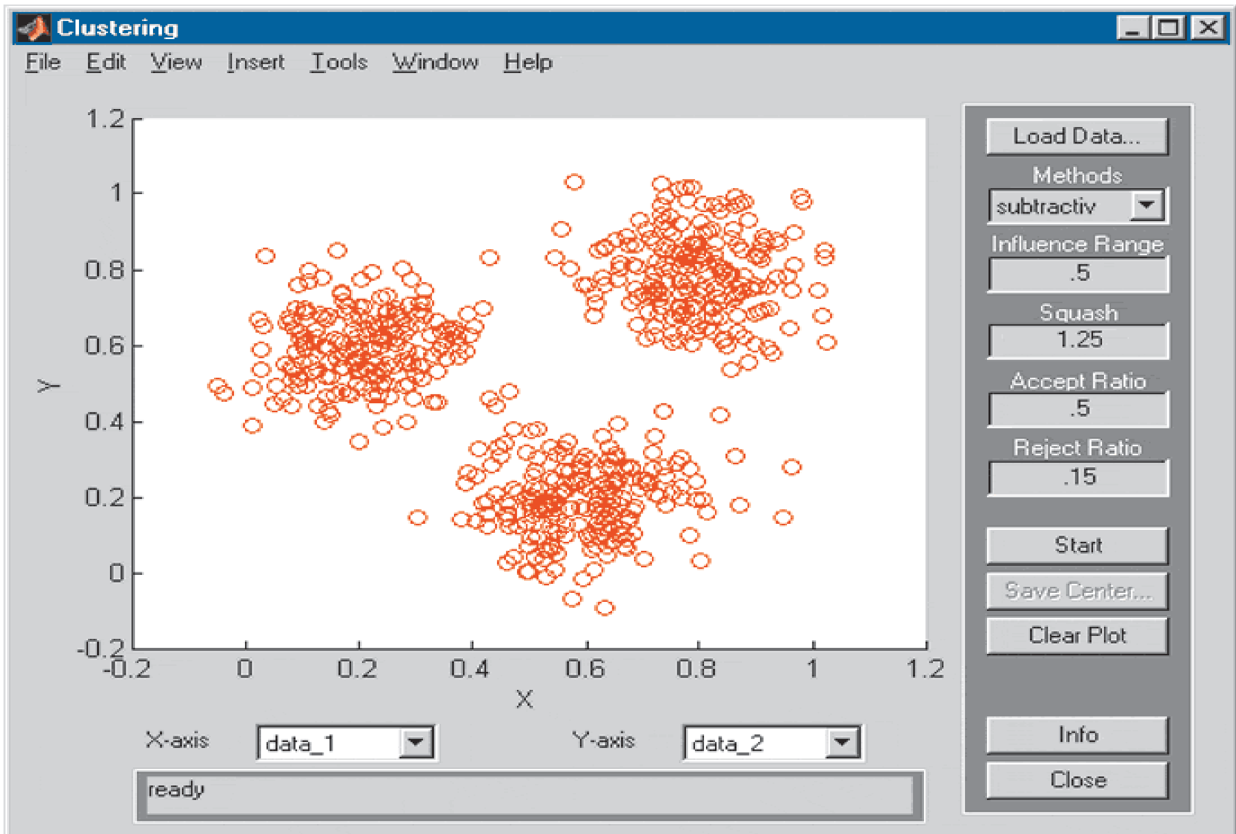




Curve fitting (regression)



Clustering (unsupervised)



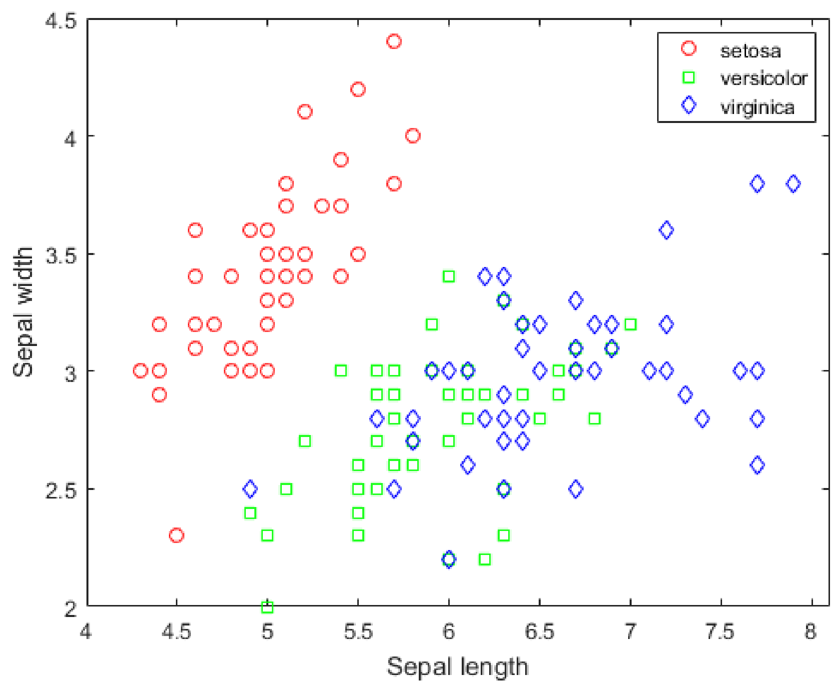
UTSA

Processes of Machine Learning

- Feature extraction and selection:
- Modeling
  - Generative model: (Bayesian based)
  - Discriminative model: (Logistic, SVM, deep learning)
  - Linear and nonlinear model
- Training
  - Infer the model parameters and structure to fit the training data
  - Bayesian and likelihood approach
- Prediction

Example: **Fisher's Iris Data**

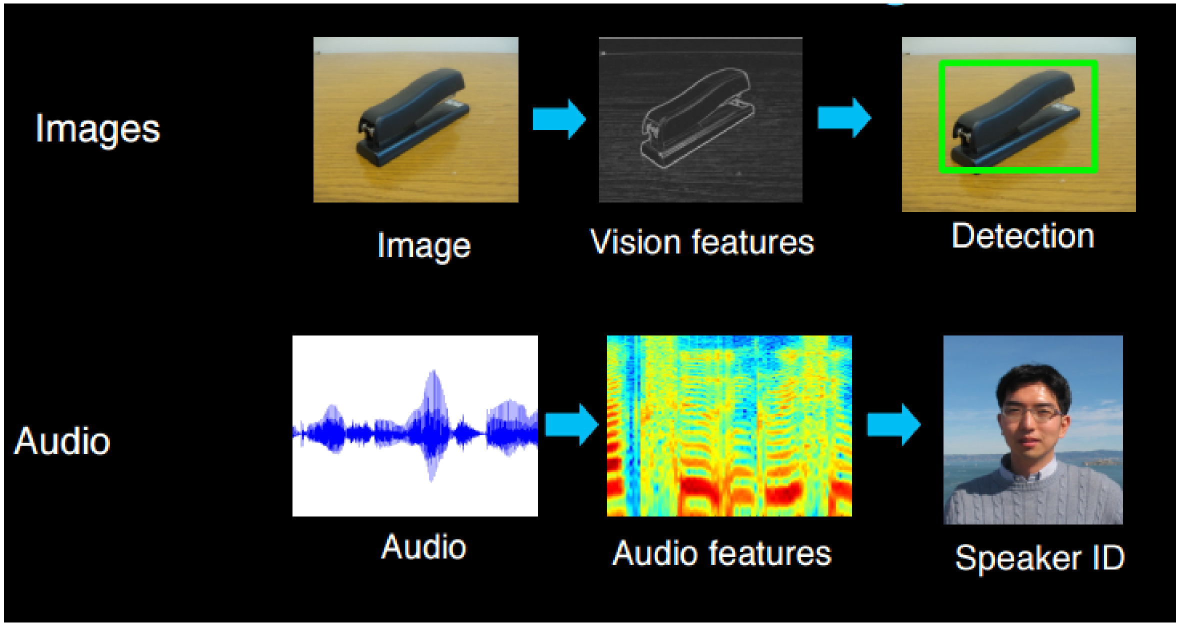
- measurements on the *sepal length*, *sepal width*, *petal* length, and *petal* width of iris samples
- three species (setosa, versicolor, virginica).
- 50 specimens from each of three species (150 iris specimens in total)



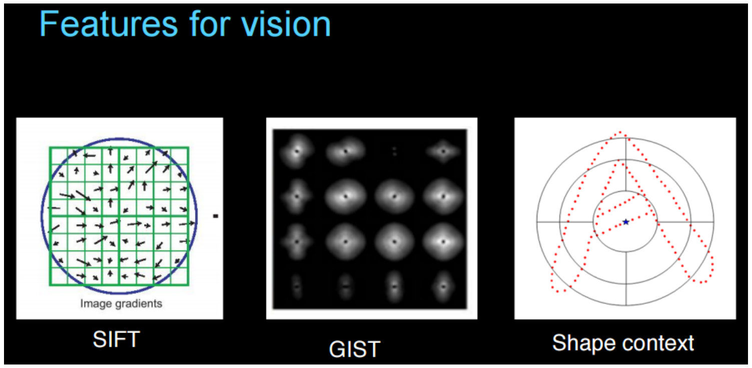
Features in Machine Learning

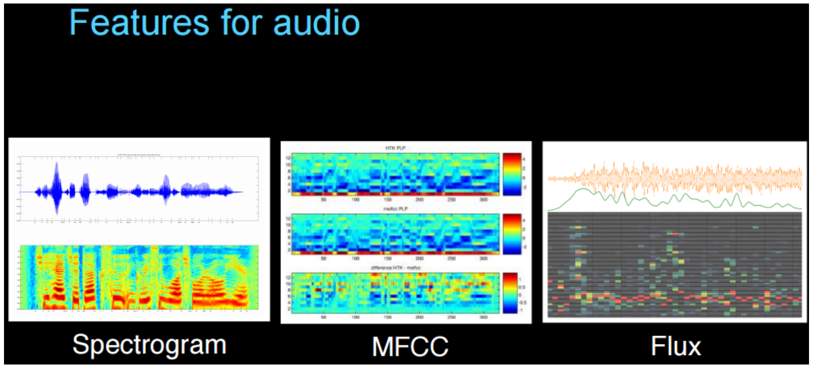
- Features are the observations that are used to form predictions
  - For image classification, the pixels are the features
  - For voice recognition, the pitch and volume of the sound samples are the features
  - For autonomous cars, data from the cameras, range sensors, and GPS are features
- Extracting relevant features is important for building a model
  - Time of day is an irrelevant feature when classifying images
  - Time of day is relevant when classifying emails because SPAM often occurs at night
- Common Types of Features in Robotics
  - Pixels (RGB data)
  - Depth data (sonar, laser rangefinders)
  - Movement (encoder values)
  - Orientation or Acceleration (Gyroscope, Accelerometer, Compass)

Examples of features



Examples of features





Feature vector

$y \in \{0, 1, 2\}$

A number of features

$$x_1, \ldots, x_l,$$

constitute the feature vector

$$\underline{x} = [x_1, \ldots, x_l]^T \in R^l$$

$\underline{x}^T = [x_1, x_2] \in R^2$

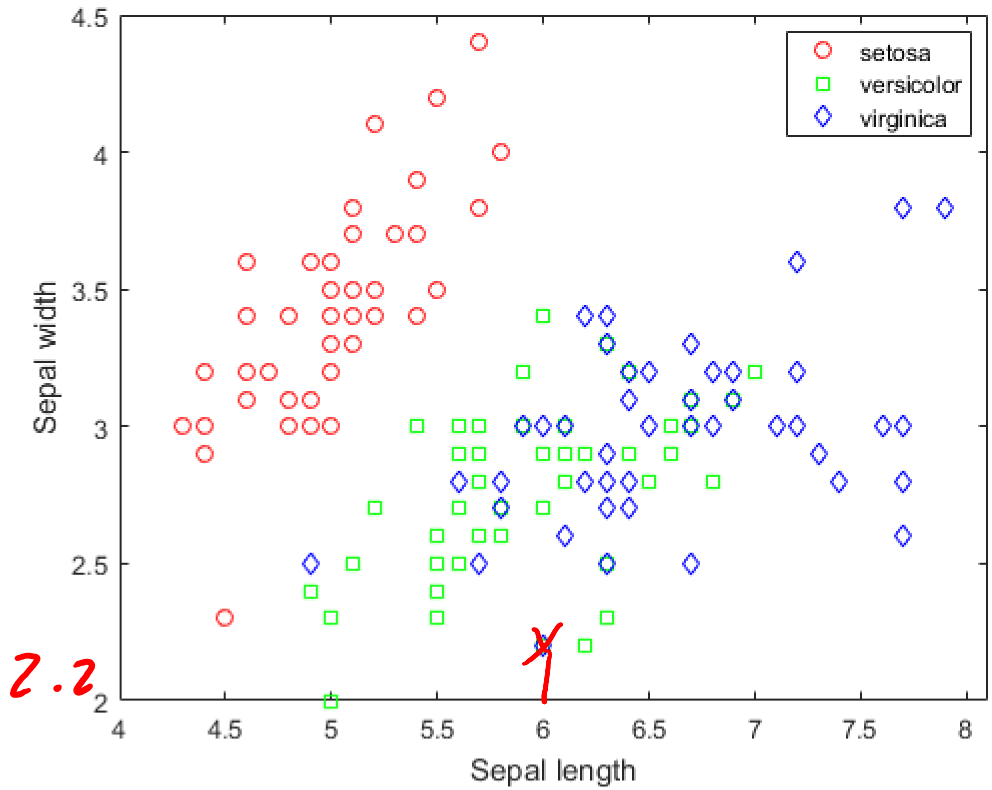
Feature vectors are treated as random vectors.

- l is called feature dimension. : 2

sample size : 150

Example: **Fisher's Iris Data**

Fisher's iris data consists of measurements on the sepal length, sepal width, petal length, and petal width for 150 iris specimens. There are 50 specimens from each of three species.



Processes of Machine Learning

- Feature extraction and selection:
- Modeling $y = f(x)$
  - Generative model: (Bayesian based)
  - Discriminative model: (Logistic, SVM, deep learning)
  - Linear and nonlinear model
- Training
  - Infer the model parameters and structure to fit the training data
  - Bayesian and likelihood approach
- Prediction

Classifier

$y \in \{0, 1 \;.\; \}$

Assign the feature vector

$$\underline{x} = \left[x_1, x_2, \ldots, x_l\right]^T$$

to the most probable of the available class labels y, that is (for binary classification)

$$\hat{y} = \begin{cases} 0 & f(\boldsymbol{x}; \boldsymbol{w}) \le 0 \\ 1 & f(\boldsymbol{x}; \boldsymbol{w}) > 0 \end{cases}$$

where

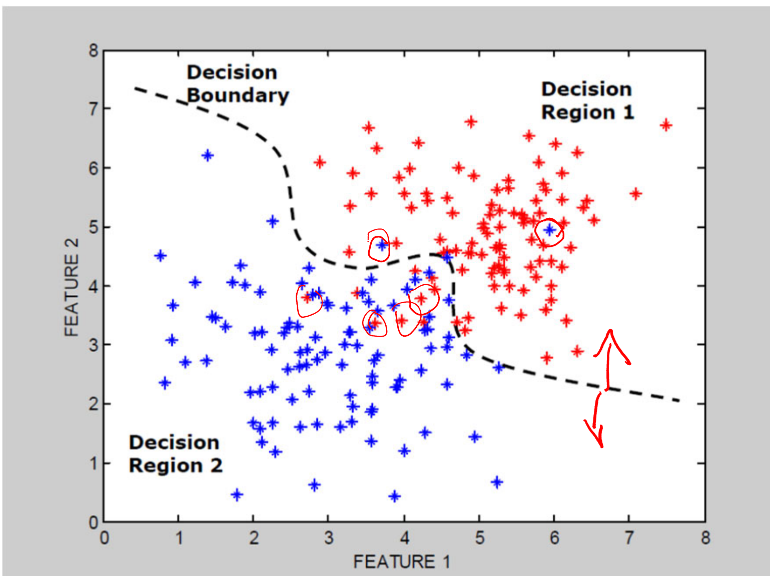
- $f(\cdot; \boldsymbol{w})$ is the classifier or discriminant function,
- $\boldsymbol{w}$ is the model parameters (to be inferred)
- Decision boundary

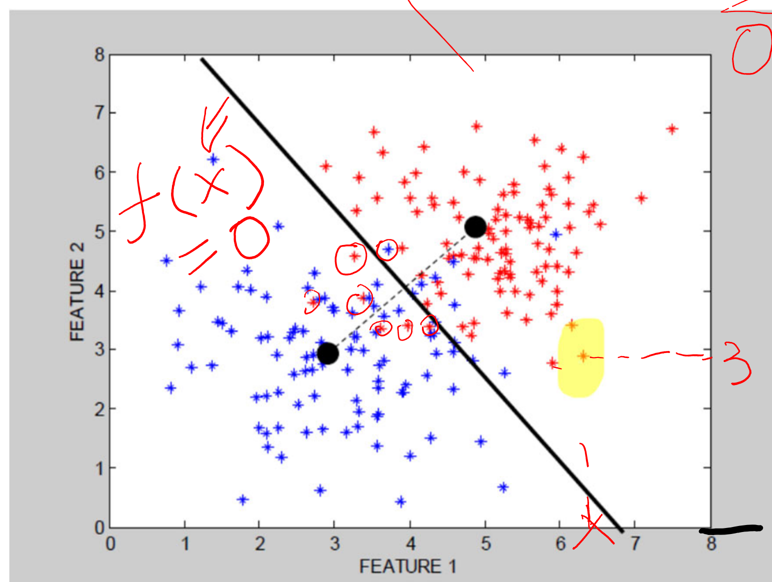
$$f(\boldsymbol{x}; \boldsymbol{w}) = 0$$

Decision boundary

Nonlinear $f(\boldsymbol{x};\boldsymbol{w}) = e^{x^2}$

Linear $f(\boldsymbol{x};\boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 \gtreqless 0$





UTSA